Dist: A

AD-A285 352

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL OR ANNUAL

4. TITLE AND SUBTITLE
Digital Adaptive and Optimal Control of Distributed Systems

5. FUNDING NUMBERS
AFOSR-91-0016
61102F 2304/AS

6. AUTHOR(S)
J.S. Gibson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90024-1406

8. PERFORMING ORGANIZATION REPORT NUMBER
Final Technical
AEOSR-TR- 94 0584

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR NM
110 DUNCAN AVE SUITE B115
BOLLING AFB DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
91-0016

11. SUPPLEMENTARY NOTES

OCT 04 1994

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
Distribution unlimited

A

12b. DISTRIBUTION CODE
UL

94-31469

13. ABSTRACT (Maximum 200 words)

Research on optimal and adaptive control and adaptive identification of distributed systems has been performed. Most of the research has focused on digital control and identification methods, to allow for real-time implementation. The main applications have been to identification and control of flexible structures.

Both new mathematical theory and new numerical methods have been primary objectives and results of the research. Experimental application of the new methods for adaptive identification and disturbance rejection has been carried out.

DTIC QUALITY INSPECTED 2

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

FINAL TECHNICAL REPORT

AFOSR GRANT 91-0016
Digital Adaptive and Optimal Control of Distributed Systems

October 1, 1990–December 31, 1993

J.S. Gibson
Mechanical, Aerospace and Nuclear Engineering
University of California, Los Angeles 90024-1597

1 Summary of Objectives and Results

This research has concerned optimal and adaptive control and adaptive identification of distributed systems. Most of the research has focused on digital control and identification methods, to allow for real-time implementation.

Both new mathematical theory and new numerical methods have been primary objectives and results of the research. Most of the new optimal control theory has concerned optimal control of distributed systems. The adaptive control methods had concentrated on adaptive disturbance rejection, or noise cancellation. The adaptive identification research has contributed both novel mathematical theory for parameter estimation and new fast numerical algorithms for adaptive filtering.

The methods for adaptive identification and disturbance rejection have been demonstrated by experimental application at Wright Patterson AFB and the Jet Propulsion Laboratory in Pasadena, CA.

2 Optimal Control of Distributed Systems

The research on optimal control of distributed parameter systems covered two basic areas: numerical methods and convergence theory for optimal control of flexible structures [6, 7, 8], and integrated design of controllers and distributed sensors for smart structures [3, 10, 18].

2.1 Numerical Methods and Convergence Theory

The research in numerical methods and convergence analysis in [6, 7, 8] continued work begun under a previous AFOSR grant. This work developed approximation methods for design of finite-dimensional compensators based on distributed models of highly flexible structures such as space antennae and satellites with large flexible solar arrays. The main results of the approximation theory were various necessary and sufficient conditions for a finite-dimensional compensator to converge to an optimal compensator for the infinite-dimensional model as the order of the compensator increases.

2.2 Integrated Design of Controllers and Distributed Sensors

The goal of the research in [3, 10, 18] is to guide the integrated design of "smart structures" and low-order compensators for smart structures. In the approach taken in [6, 7, 8], the order of a near-optimal compensator might be very large for a highly flexible structure. This drawback motivated the research in [3, 10, 18], which was initiated under this grant. This new approach leads to very low-order compensators (sometimes of dimension 1) when certain distributed sensors are designed to measure the correct functionals of the distributed state vector. While the work in [3, 10, 18] emphasizes fiber optic sensors, the same approach applies when piezo-electric films are used as distributed strain gauges.

Whether fiber-optic or piezo-electric sensors are used, the methods developed in [3, 10, 18] can be used to design flexible components for smart structures, and this was the primary motivation for the research. In this approach to the design of smart structures, the distributed sensors are embedded in the flexible beams and rods that are connected to

construct a large complex structure. Embedding fiber-optic and piezo-electric sensors in composite materials is now a proven technology.

3 Adaptive Filtering and Identification

The research on adaptive filtering and identification has included derivation of fast, numerically stable algorithms for real-time parameter estimation, as well as new theoretical methods for the study of the asymptotic properties of adaptive parameter estimates and the derivation of a new generation of adaptive filters.

3.1 Adaptive Lattice Filters

As far as useful software is concerned, the most important research under this grant was the development of new lattice filters [1, 2, 16] for adaptive identification and filtering of multichannel systems. We believe that the lattice filter in [2] is the best on the market for systems with many input and/or output channels, especially when the parallel architecture developed in [2] is used.

Least-squares lattice filters are fast algorithms for adaptive identification of linear digital input/output models. In addition to efficiency, important characteristics of lattice filters are numerical stability, order-recursiveness and suitability for parallel architectures. Recently, an new unwindowed, or covariance, lattice filter was introduced in [1] for solving the problem of exact initialization. An unwindowed lattice usually achieves much faster convergence than prewindowed lattices when the data has nonzero initial conditions.

In estimating a process with p channels, most lattice filters, including the lattice in [1], require inversion of $p \times p$ covariance matrices. The unwindowed lattice filter developed in [16] and used in this paper incorporates a Gram-Schmit procedure that renders the covariance matrices diagonal. This improves both the numerical stability and the efficiency of the lattice.

It has been demonstrated in previous literature that lattice filters can be effective for adaptive identification of flexible structures, but the previous literature used prewindowed lattices and primarily single-input/single-output models and data. The results in [11] demonstrate the advantages of an unwindowed, multi-channel lattice for identification of complex structures.

The multichannel least-squares lattice filter in [16] is used to estimate parameters for an ARX (autoregressive with exogenous input) model

$$y(t) + \sum_{i=1}^{n} A_i y(t-i) = \sum_{i=1}^{n} B_i u(t-i), \qquad t = 0, 1, 2, \ldots, \tag{3.1}$$

where $y(t)$ is an p-vector, $u(t)$ is a m-vector, and the ARMA coefficients A_i and B_i are, respectively, $p \times p$ and $p \times m$ matrices.

It should be noted that the basic residual-error lattice involves reflection coefficients rather than the ARMA coefficients A_i and B_i. The ARMA coefficients are generated from

the output of the residual error lattice by an auxiliary algorithm, which need not be run at each time step.

One of the most important features of the lattice filter developed in [16] is its suitability for VLSI realization. The forward-propagating and backward-propagating quantities on which the new lattice is based allow the lattice to be realized with wavefront arrays, which are among the most efficient VLSI architecture's. New wavefront arrays designed for the unwindowed lattice filter are presented in [16, 2].

The lattice filter in [2, 16] was used for adaptive identification of large experimental aerospace structures from input/output data in [11, 9]. The paper [11] resulted from our collaboration with Wright Patterson AFB, and [9] resulted from our collaboration with the Jet Propulsion Laboratory.

3.2 Identification of an Experimental Truss

The structure shown in Figure 1 is an experimental truss in the Flight Dynamics Laboratory at Wright Patterson Air Force Base. The truss is 12m high. Eight actuator/sensor pairs and a disturbance actuator are attached to the truss as shown. Only the actuators and sensors located on the top of the truss were used for [11].

The results reported in [11] were obtained by using the multichannel lattice filter to estimate the coefficients in (3.1) adaptively. To generate the data used for this paper, the disturbance actuator and actuator 2 excited the truss simultaneously with broad-band force sequences. These two input sequences and the corresponding output sequences from the four sensors were sampled and recorded at 50Hz, and the lattice filter was used to fit the data with ARMA models of all orders n between 1 and 40. The natural frequencies and damping ratios of the truss are estimated by computing the poles of the estimated ARMA model.

Because the geometry of the truss is almost invariant with respect to 90° rotations, the bending modes occur in pairs with almost repeated frequencies. For any one of the four sensors, certain bending modes are highly observable and certain other bending modes are marginally observable. Therefore, accurate identification of all of the modes is possible only by using all four sensors simultaneously. Figure 2 shows adaptive frequency estimates obtained using all four sensors as output channels with $n = 30$.

3.3 Adaptive Minimax Estimation and Filtering

In [4, 14, 15], we introduced a new class of parameter estimation problems, in which the estimated parameters are minimizing solutions to minimax problems for quadratic fit-to-data criteria. Whereas the asymptotic parameter estimates produced by least-squares methods are Markov parameters of Kalman filters, the asymptotic parameter estimates produced by the order-recursive minimax problem in [4] are Markov parameters of discrete-time H_∞ filters. We believe that the ideas and results in [4] constitute the theoretical foundation for a new generation of adaptive filters that will be robust to disturbances, unmodeled dynamics, and parameter variations in the plant. We are pursuing these new adaptive filters in our current research sponsored by AFOSR.

4 Adaptive Control and Disturbance Rejection

There are many aerospace and aeronautical applications of adaptive disturbance rejection, or noise cancellation. For example, the optical instruments used in aerospace telescopes typically are mounted on flexible structures in which vibrations are excited by both external disturbances and internal engines used for attitude control and other mission purposes. The optical instruments must be isolated, often actively, from such vibrations to maintain optical pathlength errors at the submicron levels required for space telemetry [9]. Also, cabin noise generated by aircraft engines and aerodynamics can be reduced significantly by active noise cancellation methods.

Under our current AFOSR grant, we have developed a new method for adaptive disturbance rejection, based on a new way of incorporating a disturbance model in an expanded plant model [17, 12, 13, 5]. While the controller design presented in [17, 13, 5] uses an internal model of the disturbance, it differs from previous disturbance-rejecting controllers based on internal disturbance models because it separates the design into two parts: design of a basic stabilizing controller for the plant, and design of a disturbance-rejecting augmentation to the basic stabilizing controller. The basic stabilizing controller for the plant is designed independently of the disturbance, and the part of the control law that stabilizes the plant is not filtered through the disturbance dynamics as in previous disturbance-rejecting controllers. Also, the disturbance-rejecting augmentation for the controller can be redesigned efficiently without changing the basic stabilizing controller. The ability to redesign the disturbance-rejecting part of the controller quickly is important for adaptive disturbance rejection in applications where the plant remains constant but the disturbance changes over time.

The method that we developed in [17, 13, 5] for adaptive disturbance rejection has performed well in simulations in [17, 5] and in an experimental application [9], where the disturbance-rejecting controller isolated a high-precision optical instrument from sinusoidal disturbances acting on a flexible truss in the JPL Phase B Test Bed (Figures 3 and 4). The test bed was developed to demonstrate sub-micron control for optical instruments to be used in future space missions. The new disturbance-rejecting controller was used to reduce optical path-length error when the structure was subjected to sinusoidal disturbances. Figure 4 shows open-loop and closed-loop path-length errors.

Publications

[1] S.-W. Chen and J.S. Gibson, "A new unwindowed lattice filter for RLS estimation," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2158–2165, September 1992.

[2] S.-B. Jiang and J.S. Gibson, "An unwindowed multichannel lattice filter with orthogonal channels," *IEEE Trans. Signal Processing*, Accepted.

[3] J.S. Gibson and C.-L. Meng, "Optimal design of fiber optic sensors for control of flexible structures," *Journal of Smart Materials and Structures*, Accepted.

[4] J.S. Gibson and G.H. Lee, "Least-squares and minimax paremeter estimation for linear systems in the presence of noise," *submitted.*

[5] W.-J. Li and J.S. Gibson, "Adaptive identification and disturbance rejection for flexible structures," *submitted.*

[6] J.S. Gibson and A. Adamian, "Approximation theory for LQG optimal control of flexible structures," *SIAM Journal on Control and Optimization*, vol. 29, pp. 1–37, January 1991.

[7] J.S. Gibson and A. Adamian, "A comparison of three approximation schemes for optimal control of a flexible structure," in *Control and Estimation in Distributed Parameter Systems*, (H. Banks, ed.), Philadelphia: SIAM, 1992.

[8] J.S. Gibson, I.G. Rosen, and G. Tao, "Approximation in Control of Thermoelastic Systems," *SIAM Journal on Control and Optimization, vol. 30, pp. 1163–1189*, September 1992.

[9] D.B. Eldred, J.S. Gibson, and W.-J. Li, "Identification and active disturbance rejection for the JPL Phase B Test Bed," in *North American Conference on Smart Structures and Materials*, (Albuquerque, New Mexico), January 1993.

[10] J.S. Gibson and C.-L. Meng, "Design of fiber optic sensors for control of flexible structures," in *American Control Conference*, (Chicago), pp. 1686–1689, IEEE, June 1992.

[11] S.-B. Jiang, J.S. Gibson, and J.J. Hollkamp, "Identification of a flexible structure by a new multichannel lattice filter," in *American Control Conference*, (Chicago), pp. 1681–1685, IEEE, June 1992.

[12] W.-J. Li and J.S. Gibson, "Variable-order adaptive identification and control of flexible structures," in *Conference on Decision and Control*, (Tucson, Az), IEEE, December 1992.

[13] W.-J. Li and J.S. Gibson, "Adaptive identification and disturbance rejection for flexible structures," in *American Control Conference*, (San Francisco), IEEE, June 1993.

[14] J.S. Gibson and G.H. Lee, "Least-squares estimation of linear systems in the presence of noise," in *Conference on Decision and Control*, (San Antonio, Tx), IEEE, December 1993.

[15] J.S. Gibson and G.H. Lee, "Minimax estimation of linear systems in the presence of noise," in *American Control Conference*, (Baltimore), IEEE, June 1994.

[16] S.-B. Jiang, *Unwindowed Multichannel Lattice Filters and Applications in Adaptive Identification and Control*. PhD thesis, University of California, Los Angeles, 1992.

[17] W.-J. Li, *Variable-order Adaptive Identification and Control of Flexible Structures with Disturbance Rejection*. PhD thesis, University of California, Los Angeles, 1992.

[18] C.-L. Meng, *Identification and control of mechanical and optical systems using distributed sensing*. PhD thesis, University of California, Los Angeles, 1993.

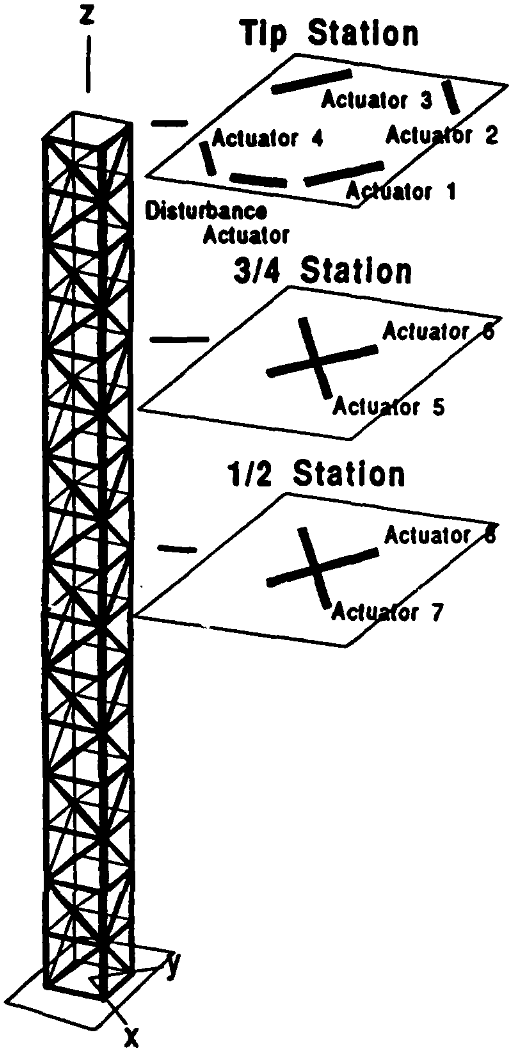

Figure 1: Wright Patterson 12-meter Truss

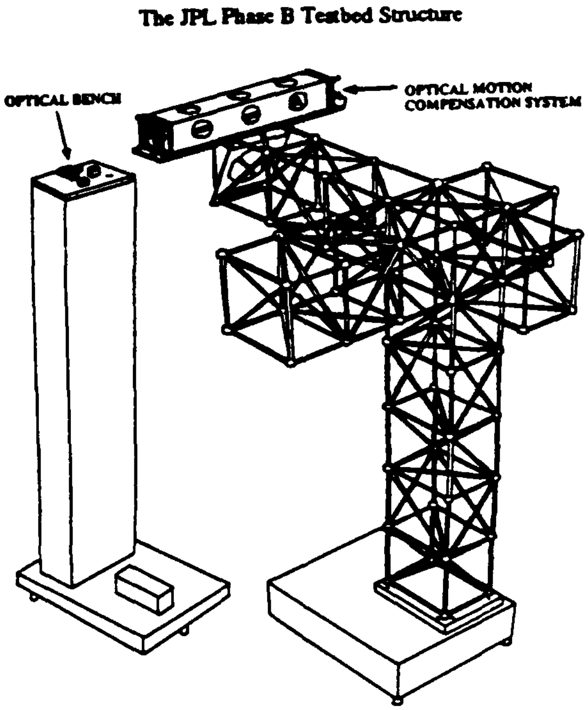

Figure 3: JPL Phase B Test Bed

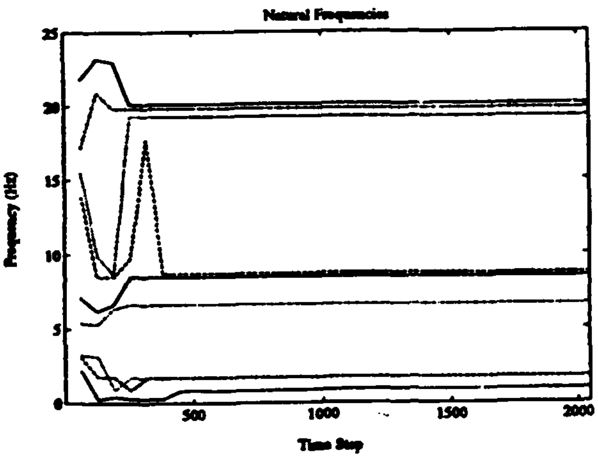

Figure 2: Nine Modes Closest to Unit Circle

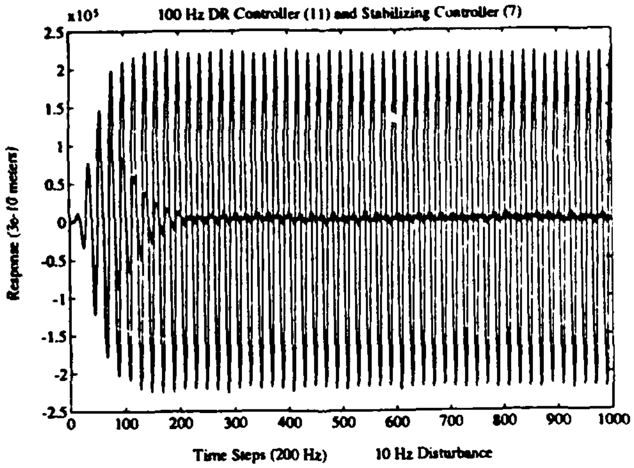

Figure 4: Controlled Output

Appendix

Publications [3] and [4]

All other publications have been provided separately.

Optimal Design of Fiber Optic Sensors for Control of Flexible Structures *

C.-L. Meng and J.S. Gibson
Mechanical, Aerospace, and Nuclear Engineering
University of California, Los Angeles 90024-1597

June 23, 1994

Abstract

This paper presents methods for designing fiber optic sensors for control of flexible structures. The sensors are designed so that a first-order observer, instead of a high-order state estimator, can be used to construct feedback control laws based on distributed or high-order finite-dimensional models of the structure. Design methods for the sensors and the observers are developed for two types of problems: first, where the sensors extend over the entire structure, and second, where the sensors extend over one or more subregions.

*This research was supported by AFOSR Grant 910016.

1 Introduction

Active control systems often are required to reduce vibrations in flexible space structures. Recent research [2], [3], [4], [5], [9], [10], [11], [12], [16], [17] has shown that fiber optic sensors can be embedded in composite materials to measure spatial integrals of strain along the length of the fiber and that such measurements can be used in feedback control systems to damp structural vibrations actively. Modal domain fiber optic sensors, discussed in [3], appear to be best suited for such applications. As shown in [11], the output of a fiber optic sensor can be represented by a weighted spatial integral of strain. Also [11] showed that, for a single actuator, a first-order dynamic observer can be used in the feedback loop to construct an arbitrary linear control law based on any finite number of structural modes. The first-order observer uses the output of two fiber optic sensors that must be tapered appropriately.

According to distributed-system theory, optimal control laws for flexible structures usually require weighted spatial integrals of strain and velocity [6]. Constructing such control laws from the measurements of local positions, velocities, accelerations, and strains using conventional point sensors requires infinite dimensional state estimators. These estimators are approximated by high-order finite dimensional estimators in implementation. Section 2 of this paper presents the design theory for a first-order functional observer that constructs control laws derived from distributed system theory. This observer design in general requires the sensors to be distributed over the entire length of the structure. Numerical examples in Section 2.4 illustrate the theoretical results presented in Section 2.1. The results in Section 2 were developed in [8, 13].

In applications, it may not be possible for the fiber optic sensors to be distributed over the entire length of a structure. When the fiber optic sensors extend only over portions of a flexible structure there will exist some error between the output of the first order functional observer and the control law that the observer is designed to construct. In this case, the design problem is to find the optimal sensor gains so that the error is minimized and the control system can achieve high performance. Section 3 discusses the design method for fiber optic sensors and the corresponding observers when the sensors extend over a portion of the structure. Section 3.3 presents the numerical examples. The results in Section 3 were developed in [13].

2 Sensors Spanning the Entire Structure

2.1 Plant Model and Energy Space

For the forced linear vibrations of flexible structures, the common abstract evolution equation is [1, 6, 7, 15]

$$\ddot{w}(t) + D_0\dot{w}(t) + A_0 w(t) = B_0 u(t). \tag{2.1}$$

The generalized displacement $w(t)$ and B_0 are in the real Hilbert space H. The forcing function $u(t)$ is a real valued function of t. The stiffness operator A_0 is a coercive self-adjoint linear operator with domain dense in H. The damping operator D_0 is a nonnegative linear operator bounded relative to A_0.

The natural strain energy space is V=Dom($A_0^{1/2}$) with inner product

$$\langle A_0^{1/2}\cdot, A_0^{1/2}\cdot\rangle_H. \tag{2.2}$$

The natural state space for first order form of (2.1) is the total energy space

$$E = V \times H. \tag{2.3}$$

2.2 A First-order Functional Observer

Theorem 1 *Let $f_1 \in V$ and $f_2 \in H$, $w(t)$ be the solution of Eq.(2.1) and define*

$$y = \langle f_1, w(t)\rangle_V + \langle f_2, \dot{w}(t)\rangle_H. \tag{2.4}$$

Let $k_1 \in V$ and $k_2 \in V$. The two measurements y_1 and y_2 are defined as:

$$\begin{aligned} y_1(t) &= \langle k_1, w(t)\rangle_V \\ y_2(t) &= \langle k_2, w(t)\rangle_V. \end{aligned} \tag{2.5}$$

For any $\beta \in R$, the real-valued function $z(t)$ is the solution of the following equation:

$$\dot{z}(t) = -z(t) + y_1 + \beta u(t). \tag{2.6}$$

Define

$$\epsilon = z(t) + y_2 - y. \tag{2.7}$$

Then ϵ satisfies:

$$\dot{\epsilon} + \epsilon = 0 \tag{2.8}$$

if and only if

$$f_2 \in V \tag{2.9}$$

$$k_2 = f_1 + (A_0^{-1} - A_0^{-1} D_0^*) f_2 \tag{2.10}$$

$$k_1 + k_2 = f_1 - f_2 \tag{2.11}$$

$$\beta = \langle f_2, B_0\rangle_H. \tag{2.12}$$

Proof Conditions (2.10)–(2.12) follow from writing the equation $\dot{\epsilon} = -\epsilon$ in detail and substituting the V-inner product in the terms involving y_1 and y_2. The inner product in the strain- energy space V is

$$\langle \cdot\,, \cdot\,\rangle_V = \langle A_0^{1/2}\,\cdot\,, A_0^{1/2}\,\cdot\,\rangle_H. \tag{2.13}$$

The condition (2.9) comes from the domain requirements of the operators A_0, $A_0^{1/2}$, D_0 and their adjoints.

In applications, $\langle f_2, \dot{w}(t)\rangle_H$ in (2.4) can be represented as

$$\langle f_2, \dot{w}(t)\rangle_H = \int_0^l f_2 \dot{w}(t) dx. \tag{2.14}$$

Measurements in (2.5) are spatial integrals of strain:

$$y_1 = \int_0^l k_1'' w''(t) dx \tag{2.15}$$

$$y_2 = \int_0^l k_2'' w''(t) dx. \tag{2.16}$$

These last two measurements can be made by fiber optic sensors. Theorem 1 says that an exponentially convergent first-order observer can be used to construct a bounded linear functional of the distributed state vector $(w, \dot{w})$ from fiber optic sensor data if and only if the velocity weighting f_2 is in the strain-energy space V. According to condition (2.9), f_2 must satisfy smoothness and boundary conditions for functions in V. These conditions will require a certain number of L_2 derivatives and that the generalized displacement function $w(t)$ satisfy certain geometric boundary conditions.

In applications, (2.10)-(2.12) must be computed numerically with finite element or modal approximation schemes. Such numerical methods project the infinite dimensional problem from the energy space E onto a sequence of finite dimensional subspaces $E_n = V_n \times V_n$, where $V_n \in V$. The operators A_0 and D_0 are approximated by A_{0n} and D_{0n} in V_n. The weighting functions f_1 and f_2 are approximated by f_{1n} and f_{2n} in V_n. The functional sensor gains k_1 and k_2 are approximated by k_{1n} and k_{2n} in V_n which satisfy

$$k_{2n} = f_{1n} + (A_{0n}^{-1} - A_{0n}^{-1} D_{0n}^{*}) f_{2n} \tag{2.17}$$

$$k_{1n} + k_{2n} = f_{1n} - f_{2n}. \tag{2.18}$$

According to [6], f_{1n} and f_{2n} will converge in V−norm to the functional control gains f_1 and f_2, respectively, in the infinite dimensional control problem. Also k_{1n} and k_{2n} will converge in V to the functional sensor gains k_1 and k_2, respectively.

2.3 Finite Dimensional Model

Consider the lateral vibration of a simply-supported uniform Euler-Bernoulli beam. The finite dimensional modal approximation scheme is used. The space can be spanned by $\{\phi_1, \phi_2, \cdots, \phi_n\}$, where ϕ_i is the ith mode shape, an $V_n = H_n$. If $w(t)$ is the solution of (2.1) and satisfies the boundary conditions. $w(t)$ can be approximated as

$$w = \sum_{i=1}^{n} \eta_i(t)\phi_i \tag{2.19}$$

where $\eta_i(t)$ is the modal coordinate of the ith mode.

Defined the state to be

$$\mathbf{x} = \begin{bmatrix} \mathbf{q} \\ \dot{\mathbf{q}} \end{bmatrix} \tag{2.20}$$

where $\mathbf{q}(t)$ is

$$\mathbf{q}(t) = \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \\ \vdots \\ \eta_n(t) \end{bmatrix}. \tag{2.21}$$

The state equation is

$$\dot{\mathbf{x}} = A\mathbf{x} + Bu \tag{2.22}$$

where

$$A = \begin{bmatrix} 0 & I \\ -A_{0n} & -D_{0n} \end{bmatrix} \tag{2.23}$$

$$B = \begin{bmatrix} 0 \\ b_{0n} \end{bmatrix}. \tag{2.24}$$

The stiffness matrix A_{0n} is:

$$A_{0n} = \begin{bmatrix} \omega_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \omega_n^2 \end{bmatrix} \tag{2.25}$$

where ω_i is the natural undamped frequency of the ith mode. There are two kinds of damping matrix D_{0n}: $c_0 I$ and $c_0 A_{0n}^{1/2}$. The choice of D_{0n} will influence the convergence of sensor gains.

The matrix b_{0n} is an $n \times 1$ column vector whose elements are the projections of the actuator function $B_0(x)$ onto mode shapes.

Suppose two fiber optic sensors extend over the entire length of a beam. Since k_{1n} and k_{2n} are in V_n, k_{1n} and k_{2n} can be represented as linear combinations of eigenfunctions:

$$k_{1n}(x) = \sum_1^n \xi_j \phi_j(x) \tag{2.26}$$

$$k_{2n}(x) = \sum_1^n \sigma_j \phi_j(x). \tag{2.27}$$

The two measurements y_1 and y_2 are

$$y_1(t) = \langle k_{1n}, w(t)\rangle_{V_n} = \int_0^l k''_{1n} w''(t) dx \tag{2.28}$$

$$y_2(t) = \langle k_{2n}, w(t)\rangle_{V_n} = \int_0^l k''_{2n} w''(t) dx. \tag{2.29}$$

Define

$$\mathbf{k_{1n}} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \quad \mathbf{k_{2n}} = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_n \end{bmatrix}. \tag{2.30}$$

(2.28) and (2.29) are equal to:

$$y_1(t) = \mathbf{k_{1n}}^T A_{0n} \mathbf{q} \tag{2.31}$$

$$y_2(t) = \mathbf{k_{2n}}^T A_{0n} \mathbf{q}. \tag{2.32}$$

Let $f_{1n} \in V_n$ and $f_{2n} \in V_n$, and

$$\begin{aligned} y &= \langle f_{1n}, w(t)\rangle_{V_n} + \langle f_{2n}, \dot{w}(t)\rangle_{H_n} \\ &= \mathbf{f_{1n}}^T A_{0n} \mathbf{q} + \mathbf{f_{2n}}^T \dot{\mathbf{q}} \end{aligned} \tag{2.33}$$

where

$$\mathbf{f_{1n}} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix} \quad \mathbf{f_{1n}} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} \tag{2.34}$$

and γ_i and δ_i are the Fourier coefficients of f_{1n} and f_{2n} on ith mode shape respectively.

The first order functional observer for this model is

$$\dot{z}(t) = -z + y_1 + \beta u(t). \tag{2.35}$$

Define

$$\epsilon = z(t) + y_2 - y \tag{2.36}$$

and ϵ satisfies

$$\dot{\epsilon} = \epsilon. \tag{2.37}$$

Substituting (2.31), (2.32), (2.33) and (2.36) into (2.37), one can have the following matrix equations

$$\mathbf{k_{2n}} = \mathbf{f_{1n}} + (A_{0n}^{-1} - A_{0n}^{-1} D_{0n}^T)\mathbf{f_{2n}} \tag{2.38}$$

$$\mathbf{k_{1n}} + \mathbf{k_{2n}} = \mathbf{f_{1n}} - \mathbf{f_{2n}} \tag{2.39}$$

$$\beta = \mathbf{f_{2n}}^T b_{0n}. \tag{2.40}$$

So $\mathbf{k_{1n}}$, $\mathbf{k_{2n}}$ can be solved.

2.4 Examples

Consider the simply supported Euler-Bernoulli beam. The real Hilbert space $H = L_2(0,1)$, $V = H_0^2(0,1)$. The stiffness operator A_0 is

$$A_0\phi = \phi''''. \tag{2.41}$$

The actuator function is

$$B_0(x) = \begin{cases} \sin 5\pi(x-0.3) & \text{if } 0.3 \le x \le 0.5 \\ 0 & \text{otherwise.} \end{cases} \tag{2.42}$$

Two cases of damping are considered

Case 1: $D_0 = c_0 A_0^{1/2}$ with $c_0 = 0.02$

Case 2: $D_0 = c_0 I$ with $c_0 = 0.02\pi^2$.

Figures 2.1-2.8 are generated using the finite dimensional model describes in Section 2.3, assuming $EI = 1$, $m = 1$. Also $D_{0n} = c_0 A_{0n}^{1/2}$ in Case 1 and $D_{0n} = c_0 I$ in Case 2.

2.4.1 Constant Velocity Weighting: A Counter Example

Suppose we want to build a first order functional observer to construct a spatial integral of transverse velocity over a beam from two fiber optic sensors data. Let $f_1 = 0$ and $f_2 = 1$ in (2.4). For a simply-supported beam, two of the geometric boundary conditions for functions in V are zero at both ends, so $f_2 \notin V$. From Theorem 1, there exist no functional sensor gains k_1 and k_2 such that the desired $y(t)$ can be constructed.

In finite dimensional approximation, the vector $\mathbf{f_{1n}}$ is a zero vector and the vector $\mathbf{f_{2n}}$ is the L_2-projection of $f_2 = 1$ onto V_n. Then solving (2.39) and (2.38), one can get $\mathbf{k_{1n}}$ and $\mathbf{k_{2n}}$, which are the modal coefficients vectors of k_{1n} and k_{2n}. A first order observer can be built by using two fiber optic sensors designed according to k_{1n} and k_{2n} and the observer can construct the integral over the beam of the projection of velocity $\dot{w}(t)$ onto V_n.

But as n increases, k_{1n} and k_{2n} do not converge in V. As a result, k_1'' and k_2'' diverge in L_2. Figures 2.1 and 2.2 illustrate the divergence of k_1'' in both damping case.

2.4.2 An Optimal Control Problem

For both damping cases, let f_1 and f_2 be the functional control gains which minimize the performance index:

$$J = \int_0^\infty (\|w\|_V^2 + \|\dot{w}\|_H^2 + u^2)dt. \tag{2.43}$$

In finite dimensional approximation, f_{1n} and f_{2n} can be obtained by solving the Riccati Equation. [6] guarantees that f_{1n} and f_{2n} will converge to f_1 and f_2 in V and H, respectively. Also k_1 and k_2 are approximated by k_{1n} and k_{2n} and can be obtained by solving (2.39) and (2.38). The numerical results show that k_2'' is two orders of magnitude smaller than k_1''. Therefore we conclude that measurement of y_2 can be omitted.

Figures 2.3-2.8 show f_{1n}'', f_{2n} and k_{1n}'' with n increasing for two damping cases. For $D_0 = c_0 A_0^{1/2}$, Figure 2.4 shows that f_{2n} converges to f_2 and $f_2 \in V$. So condition (2.9) is satisfied and k_1, k_2 exist. Figure 2.5 shows that k_{1n}'' converges to k_1''.

For the case $D_0 = c_0 I$, Figure 2.7 shows that f_{2n} still converges to f_2 with $f_2 \in H$, but $f_2 \notin V$. Because at $x = 0.3$ and $x = 0.5$, f_2 has discontinuous derivative. This fact can be verified by the plots for large n. So k_1 and k_2 do not exist according to the theorem. Figure 2.8 shows that k_{1n}'' diverges as n increases. Therefore k_1'' diverges.

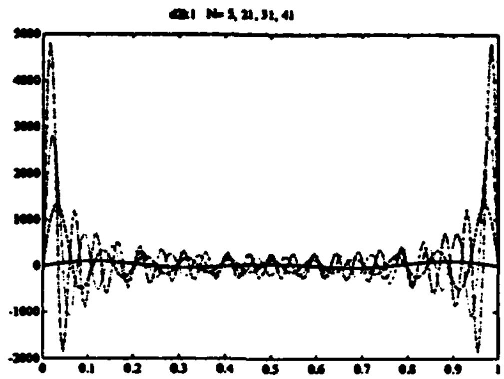

Figure 2.1: k''_{1n} for $f_2 = 1$, $D_0 = c_0 I$, $n = 5, 21, 31, 41$

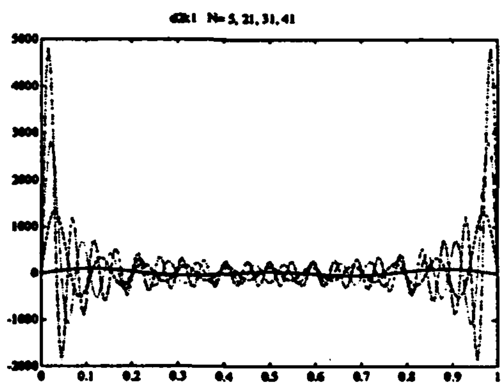

Figure 2.2: k''_{1n} for $f_2 = 1$, $D_0 = c_0 A_0^{1/2}$, $n = 5, 21, 31, 41$

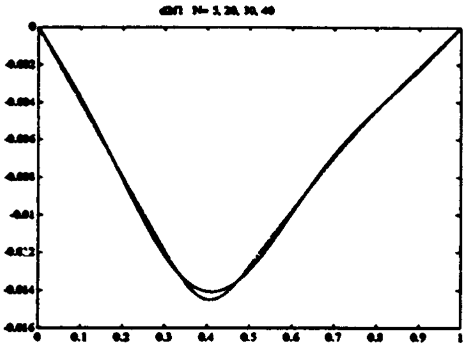

Figure 2.3: f_{1n}'', $D_0 = c_0 A_0^{1/2}$, $n = 5, 20, 30, 40$

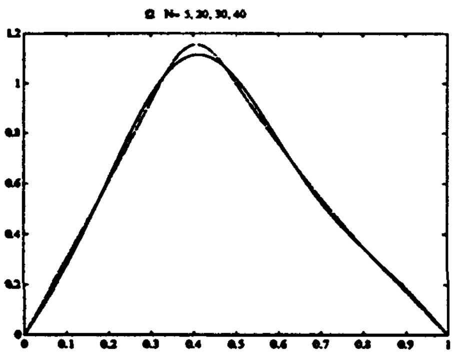

Figure 2.4: f_{2n}, $D_0 = c_0 A_0^{1/2}$, $n = 5, 20, 30, 40$

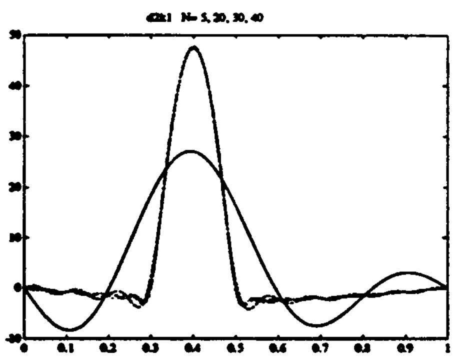

Figure 2.5: k_{1n}'', $D_0 = c_0 A_0^{1/2}$, $n = 5, 20, 30, 40$

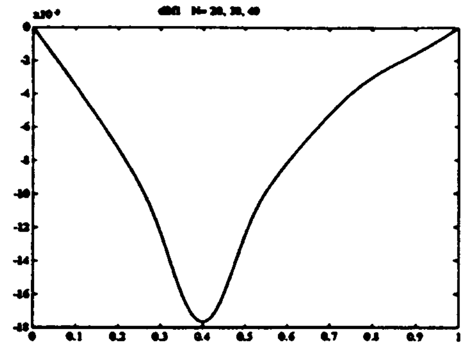

Figure 2.6: f_{1n}'', $D_0 = c_0 I$, $n = 20, 30, 40$

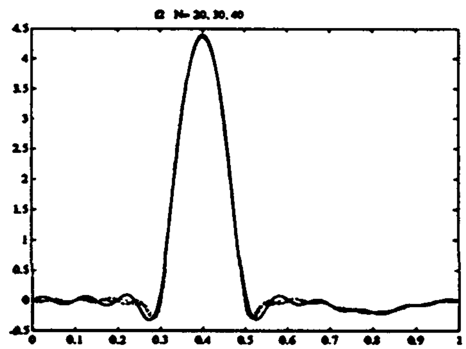

Figure 2.7: f_{2n}, $D_0 = c_0 I$, $n = 20, 30, 40$

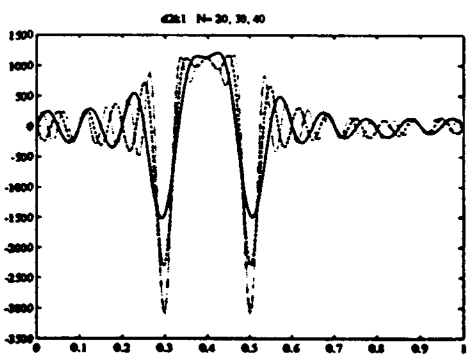

Figure 2.8: k_{1n}'', $D_0 = c_0 I$, $n = 20, 30, 40$

3 Optimization of Sensors Designs for Limited Sensor Spans

3.1 Sensors Extending Over Subintervals of a Beam

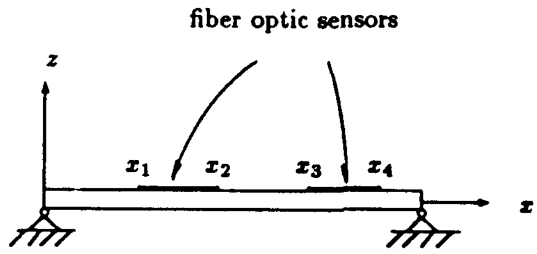

Figure 3.9: Sensors extend over portions of the beam

Suppose the first fiber optic sensor is from x_1 to x_2 and the second one is from x_3 to x_4, as shown in Fig. 3.9. The outputs of the sensors can be expressed as:

$$y_1 = \int_{x_1}^{x_2} k_1''(x) w''(x,t)\, dx \tag{3.44}$$

and

$$y_2 = \int_{x_3}^{x_4} k_2''(x) w''(x,t)\, dx \tag{3.45}$$

The finite dimensional approximation of the sensor outputs can be obtained as in the following. ϕ_i is the ith mode shape of a simply supported beam. Define $\tilde{\phi}_i$ and $\hat{\phi}_i$, $i = 1, \ldots, n$, in such a way that

$$\tilde{\phi}_i''(x) = \begin{cases} \phi_i''(x) & \text{if } x_1 \le x \le x_2 \\ 0 & \text{otherwise} \end{cases} \tag{3.46}$$

$$\hat{\phi}_i''(x) = \begin{cases} \phi_i''(x) & \text{if } x_3 \le x \le x_4 \\ 0 & \text{otherwise.} \end{cases} \tag{3.47}$$

The geometric boundary conditions are:

$$\begin{aligned} \tilde{\phi}_i(0) &= \tilde{\phi}_i(l) = 0 \\ \hat{\phi}_i(0) &= \hat{\phi}_i(l) = 0. \end{aligned} \tag{3.48}$$

There are unique solutions of $\tilde{\phi}_i$ and $\hat{\phi}_i$ for each i. k_1 and k_2 can be approximated as

$$k_1 = \sum_{j=1}^{n} k_{1j}\, \tilde{\phi}_j \tag{3.49}$$

and

$$k_2 = \sum_{j=1}^{n} k_{2j}\, \hat{\phi}_j. \tag{3.50}$$

Let $\eta_i(t)$ be the modal coordinate of the ith mode. Then the lateral displacement is approximated as

$$w = \sum_{j=1}^{n} \eta_j(t) \phi_j. \tag{3.51}$$

Substituting (3.49), (3.50) and (3.51) into (3.44) and (3.45), the finite dimensional approximation of the sensor outputs are

$$y_1 = \mathbf{k_1}^T W_1 \mathbf{q}(t) \tag{3.52}$$

$$y_2 = \mathbf{k_2}^T W_2 \mathbf{q}(t) \tag{3.53}$$

where $\mathbf{q}(t)$ is defined in (2.21) and $\mathbf{k_1}, \mathbf{k_2}$ are

$$\mathbf{k_1} = \begin{bmatrix} k_{11} \\ k_{12} \\ \vdots \\ k_{1n} \end{bmatrix} \quad \mathbf{k_2} = \begin{bmatrix} k_{21} \\ k_{22} \\ \vdots \\ k_{2n} \end{bmatrix} \tag{3.54}$$

and W_1 and W_2 are $n \times n$ matrices with elements

$$[W_1]_{ij} = \int_{x_1}^{x_2} \phi_i''(x)\phi_j''(x)dx \tag{3.55}$$

$$[W_2]_{ij} = \int_{x_3}^{x_4} \phi_i''(x)\phi_j''(x)dx. \tag{3.56}$$

Suppose we want to construct a functional observer whose output will converge to some control law asymptotically. The state equation is defined in (2.20) and (2.22). The output of the system is

$$y_1 \;=\; C\mathbf{x} \;=\; [C_1\, 0_{1\times n}]\mathbf{x} \tag{3.57}$$

where $C_1 = \mathbf{k_1}^T W_1$. The functional observer is defined as

$$\begin{cases} \dot{z} & = Fz + Gy_1 + \beta u \\ \varphi & = z + \gamma y_2. \end{cases} \tag{3.58}$$

Let $y = \mathbf{f_{1n}}^T A_{0n}\mathbf{q} + \mathbf{f_{2n}}^T \dot{\mathbf{q}}$ be the control law that the functional observer is designed to construct. Assume that $G = \gamma = 1$, $F = -1$, and define the error between the output of the functional observer and the control law to be

$$\epsilon = \varphi - y. \tag{3.59}$$

ϵ satisfies

$$\dot{\epsilon} + \epsilon = 0 \tag{3.60}$$

if and only if

$$W_2\mathbf{k_2} = A_{0n}\mathbf{f_{1n}} - D_{0n}\mathbf{f_{2n}} + \mathbf{f_{2n}} \tag{3.61}$$

$$W_1\mathbf{k_1} + A_{0n}\mathbf{f_{2n}} = -W_2\mathbf{k_2} + A_{0n}\mathbf{f_{1n}} \tag{3.62}$$

$$\beta = \mathbf{f_{2n}}^T b_{0n}. \tag{3.63}$$

For the finite dimensional model, the unique solutions of $\mathbf{k_1}$ and $\mathbf{k_2}$ can be obtained by solving the above equations if W_1 and W_2 are nonsingular. So the finite dimensional functional sensor gains k_1 and k_2 can be obtained. As the number of modes that is used in modal approximation increases, k_1 and k_2 will diverge because the sensors extend only over portions of the beam. As a result, the error will always exist.

On the other hand, if sensors are distributed over the entire length of a beam, W_1 and W_2 are both equal to the stiffness matrix A_{0n}, and the finite dimensional solutions $\mathbf{k_1}$ and $\mathbf{k_2}$ definitely exist. The functional sensor gains will converge as long as conditions in Theorem 1 are satisfied and error will go to zero asymptotically.

Suppose we allow the error to exist, the design problem now is to find the optimal functional sensor gains k_1 and k_2, which will converge and minimize the error in some sense.

3.2 Optimization of the Sensor Gains

The open-loop design procedure for the optimal functional sensor gains k_1 and k_2 is presented, so that the L_2 norm of the error ϵ between output of the functional observer and the control law will be minimized. Suppose the length of the beam is l. Define performance index to be:

$$J = \int_0^\infty \epsilon(t)^2 dt + r\int_0^l [(k_1''')^2(x) + (k_2''')^2(x)]dx \tag{3.64}$$

There are two reasons that we add the second integral to the performance index. The first reason is to make k_1'' and k_2'' smoother so that the manufacturing of fiber optic sensors is feasible. Considering (3.44) and (3.45), the outputs of the fiber optic sensors are weighted spatial integrals of strain over subintervals of a beam. Fiber optic sensors are tapered according to the value of k_1'' and k_2'' along their distribution. Therefore smoother k_1'' and k_2'' can make the sensors easier to be manufactured. Also, k_1'' and k_2'' can be interpreted as the sensitivity to local strain. If k_1'' and k_2'' are not smooth, then there may be a abrupt change in sensitivity between two infinitely close points on the beam. That will make it impossible to build the sensors.

The second reason is that penalizing $(k_1''')^2 + (k_2''')^2$ in the performance index can make k_1'' and k_2'' converge. From the discussion in Section 3.1, when the sensors extend over only subintervals of the beam, error may always exist. So if we only penalize $\int_0^\infty \epsilon(t)^2 dt$, k_1'' and k_2'' will diverge. On the other hand, just a small value of r can make them converge. Large r can make k_1'' and k_2'' smoother and converge faster but will result in a large value of open-loop error in loop gain. So it is a design trade off.

$\mathbf{k_1}$ and $\mathbf{k_2}$ are two finite dimensional column vectors whose elements are the coefficients of k_1 and k_2 on each mode, respectively. Define

$$\mathbf{k} = \begin{bmatrix} \mathbf{k_1} \\ \mathbf{k_2} \end{bmatrix} \tag{3.65}$$

The performance index can be represented as a linear-quadratic functional of $\mathbf{k}$.

3.2.1 Augmented Plant and L_2 Norm of Error

The state equation of the beam is defined in (2.22) and the first order functional observer is defined in (3.58). The augmented state is defined to be:

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ z \end{bmatrix}. \tag{3.66}$$

Then the augmented state equation becomes:

$$\dot{\tilde{\mathbf{x}}} = \tilde{A}\tilde{\mathbf{x}} + \tilde{B}u \tag{3.67}$$

where

$$\tilde{A} = \begin{bmatrix} A & 0 \\ GC & F \end{bmatrix} \tag{3.68}$$

$$\tilde{B} = \begin{bmatrix} B \\ \beta \end{bmatrix}. \tag{3.69}$$

Let $y = \mathbf{l_1}^T\mathbf{q} + \mathbf{l_2}^T\dot{\mathbf{q}}$ be the optimal control law where $\mathbf{l_1}$ and $\mathbf{l_2}$ are $n \times 1$ vectors. The measurements from the fiber optic sensors are $y_1 = \mathbf{k_1}^T W_1\mathbf{q}$ and $y_2 = \mathbf{k_2}^T W_2\mathbf{q}$, where W_1 and W_2 are defined in Section 3.1. The output of the observer is $\varphi = z + \gamma y_2$. Then the error can be written as:

$$\begin{aligned} \epsilon &= \varphi - y \\ &= \tilde{C}\tilde{\mathbf{x}} \end{aligned} \tag{3.70}$$

where

$$\tilde{C} = [\ (\gamma \mathbf{k_2}^T W_2 - \mathbf{l_1}^T) \quad -\mathbf{l_2}^T \quad 1\ ]. \tag{3.71}$$

Suppose the input $u(t)$ is an impulse function. Because the augmented system is stable, $\epsilon(t)$ is square Lebesgue integrable,

$$\int_0^\infty \epsilon(t)^2 dt < \infty \tag{3.72}$$

and $\epsilon(t) \in L_2[0,\infty)$. The norm of $\epsilon(t)$ in time domain space $L_2[0,\infty)$ is

$$\|\epsilon\|_2 = [\int_0^\infty \epsilon(t)^2 dt]^{1/2}. \tag{3.73}$$

Further assume that the initial state of the augmented system is zero. One can have

$$\epsilon(t) = \tilde{C} e^{\tilde{A}t} \tilde{B} \tag{3.74}$$

and

$$\|\epsilon\|_2^2 = tr\ \tilde{B}\tilde{B}^T \int_0^\infty e^{\tilde{A}^T t} \tilde{C}^T \tilde{C} e^{\tilde{A}t} dt. \tag{3.75}$$

Since $\tilde{A}$ is asymptotically stable, the Lyapunov equation

$$\tilde{A}^T Q + Q\tilde{A} = -\tilde{C}^T \tilde{C} \tag{3.76}$$

has exactly one solution Q for each $-\tilde{C}^T\tilde{C}$ and this solution is

$$Q = \int_0^\infty e^{\tilde{A}^T t} \tilde{C}^T \tilde{C} e^{\tilde{A}t} dt. \tag{3.77}$$

Q is symmetric and nonnegative. Therefore the square of L_2 norm of $\epsilon(t)$ is equivalent to

$$\|\epsilon(t)\|_2^2 = tr\ \tilde{B}\tilde{B}^T Q. \tag{3.78}$$

3.2.2 Frequency Domain

Let $\hat{\epsilon}(j\omega)$ be the Fourier Transform of the function $\epsilon(t) \in L_2[0,\infty)$. From the Parseval's relation for aperiodical signal,

$$\int_0^\infty \epsilon(t)^2 dt \quad = \quad \frac{1}{2\pi} \int_{-\infty}^\infty \hat{\epsilon}(j\omega)^2 d\omega \quad < \quad \infty \tag{3.79}$$

therefore $\hat{\epsilon}(j\omega)$ is square Lebesgue integrable and $\hat{\epsilon}(j\omega)$ is in the frequency domain space $L_2(-\infty,\infty)$. The L_2 norm of $\hat{\epsilon}(j\omega)$ is

$$\|\hat{\epsilon}(j\omega)\|_2 = [\frac{1}{2\pi} \int_{-\infty}^\infty \hat{\epsilon}(j\omega)^2 d\omega]^{1/2}. \tag{3.80}$$

The meaning of $\|\hat{\epsilon}(j\omega)\|_2^2$ is the energy of the signal $\hat{\epsilon}(j\omega)$. Therefore minimize the energy of signal $\hat{\epsilon}(j\omega)$ in frequency domain when the system is subject to sinusoidal input is equivalent to minimize $\|\epsilon(t)\|_2^2$ in time domain.

Under the previous assumptions of impulse input and zero initial augmented state, $\hat{\epsilon}(j\omega)$ is

$$\hat{\epsilon}(j\omega) = \tilde{C}(j\omega I - \tilde{A})^{-1} \tilde{B}. \tag{3.81}$$

3.2.3 Solving the Lyapunov Equation

In order to minimize $\|\epsilon(t)\|_2^2$ in (3.78), we have to solve (3.76) first. Partition the solution matrix Q in the following way:

$$Q = \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \tag{3.82}$$

where Q_1, Q_2, Q_3 and Q_4 are $2n \times 2n$, $2n \times 1$, $1 \times 2n$ and 1×1 matrices, respectively. Then (3.78) can be rewritten as

$$\|\epsilon(t)\|_2^2 = tr\ BB^T Q_1 + 2B^T Q_2 \beta - \frac{\beta^2}{2F}. \tag{3.83}$$

(3.76) becomes:

$$\begin{bmatrix} A^T & C^T G \\ 0^T & F \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} + \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} \begin{bmatrix} A & 0 \\ GC & F \end{bmatrix} = -\tilde{C}^T \tilde{C}. \tag{3.84}$$

After straightforward matrix multiplication, we can have the following equations:

$$A^T Q_1 + Q_1 A =$$

$$-\begin{bmatrix} (\gamma W_2 \mathbf{k_2} - \mathbf{l_1}) \\ -\mathbf{l_2} \end{bmatrix} \begin{bmatrix} (\gamma \mathbf{k_2}^T W_2 - \mathbf{l_1}^T) & -\mathbf{l_2}^T \end{bmatrix} - G(C^T Q_2^T + Q_2 C) \tag{3.85}$$

$$A^T Q_2 + C^T G Q_4 + Q_2 F = -\begin{bmatrix} (\gamma W_2 \mathbf{k_2} - \mathbf{l_1}) \\ -\mathbf{l_2} \end{bmatrix} \tag{3.86}$$

$$Q_2 = Q_3^T \tag{3.87}$$

$$Q_4 = \frac{-1}{2F} \tag{3.88}$$

where γ, G and F are all scalars. Define

$$H = \begin{bmatrix} 0 & 0 \\ W_2 & 0 \end{bmatrix} \quad D = \begin{bmatrix} W_1 & 0 \\ 0 & 0 \end{bmatrix} \quad \mathbf{l} = \begin{bmatrix} \mathbf{l_1} \\ \mathbf{l_2} \end{bmatrix}. \tag{3.89}$$

Q_2 can be simplified to be a linear function of $\mathbf{k}$:

$$\begin{aligned} Q_2 &= (A^T + FI)^{-1} \left(-\begin{bmatrix} (\gamma W_2 \mathbf{k_2} - \mathbf{l_1}) \\ -\mathbf{l_2} \end{bmatrix} + \tfrac{G}{2F} C^T \right) \\ &= (A^T + FI)^{-1} [(-\gamma H^T + \frac{G}{2F} D^T)\mathbf{k} + \mathbf{l}] \\ &= M_1 \mathbf{k} + M_2 \end{aligned} \tag{3.90}$$

where $M_1 = (A^T + FI)^{-1}(-\gamma H^T + \frac{G}{2F} D^T)$ and $M_2 = (A^T + FI)^{-1}\mathbf{l}$. (3.85) can be arranged to be a Lyapunov equation

$$A^T Q_1 + Q_1 A = -\tilde{R} \tag{3.91}$$

where $\tilde{R}$ is defined as:

$$\begin{aligned} \tilde{R} = &(\gamma^2 H^T \mathbf{k}\mathbf{k}^T H - \gamma H^T \mathbf{k}\mathbf{l}^T - \gamma \mathbf{l}\mathbf{k}^T H + \mathbf{l}\mathbf{l}^T) \\ &+ G(D^T \mathbf{k}\mathbf{k}^T M_1^T + M_1 \mathbf{k}\mathbf{k}^T D + D^T \mathbf{k} M_2^T + M_2 \mathbf{k}^T D). \end{aligned} \tag{3.92}$$

Since A is asymptotically stable, the solution of (3.91) is

$$Q_1 = \int_0^\infty e^{A^T t} \tilde{R} e^{At} dt. \tag{3.93}$$

3.2.4 Linear-Quadratic Functional of k

Consider (3.83) first. The term $tr\ BB^TQ_1$ is equal to

$$\begin{aligned} tr\ BB^TQ_1 & \\ = \ & \gamma^2\mathbf{k}^T H \int_0^\infty e^{At}BB^Te^{A^Tt}dt\ H^T\mathbf{k} \\ - \ & 2\gamma\mathbf{l}^T \int_0^\infty e^{At}BB^Te^{A^Tt}dt\ H^T\mathbf{k} \\ + \ & \mathbf{l}^T \int_0^\infty e^{At}BB^Te^{A^Tt}dt\ \mathbf{l} \\ + \ & G\mathbf{k}^TM_1^T \int_0^\infty e^{At}BB^Te^{A^Tt}dt\ D^T\mathbf{k} \\ + \ & 2GM_2^T \int_0^\infty e^{At}BB^Te^{A^Tt}dt\ D^T\mathbf{k} \\ + \ & G\mathbf{k}^TD \int_0^\infty e^{At}BB^Te^{A^Tt}dt\ M_1\mathbf{k}. \end{aligned} \tag{3.94}$$

Define

$$V = \int_0^\infty e^{At}BB^Te^{A^Tt}dt. \tag{3.95}$$

Then V is the solution of the Lyapunov equation:

$$AV + VA^T = -BB^T. \tag{3.96}$$

The second term in (3.83) is equal to

$$2B^TQ_2\beta = 2\beta B^T(M_1K + M_2). \tag{3.97}$$

Therefore square of L_2 norm of error is a linear-quadratic form of $\mathbf{k}$:

$$\begin{aligned} \int_0^\infty \epsilon(t)^2dt \ = \ & \mathbf{k}^T(\gamma^2HVH^T + GM_1^TVD + GDVM_1)\mathbf{k} \\ - \ & 2(\gamma\mathbf{l}^TVH^T - GM_2^TVD^T - \beta B^TM_1)\mathbf{k} \\ + \ & \mathbf{l}^TV\mathbf{l} + 2\beta B^TM_2 - \frac{\beta^2}{2F} \end{aligned} \tag{3.98}$$

where F, G and γ are scalars.

Consider the term $\int_0^l[(k_1''')^2(x) + (k_2''')^2(x)]dx$ in performance index.

$$\int_0^l [(k_1''')^2(\eta) + (k_2''')^2(\eta)]d\eta = \int_{x_1}^{x_2} (k_1''')^2(x)dx + \int_{x_3}^{x_4} (k_2''')^2(x)dx \tag{3.99}$$

Define U_1 and U_2 such that

$$[U_1]_{ij} = \int_{x_1}^{x_2} \phi_i'''(x)\phi_j'''(x)dx \tag{3.100}$$

$$[U_2]_{ij} = \int_{\eta_3}^{\eta_4} \phi_i'''(\eta)\phi_j'''(\eta)d\eta. \tag{3.101}$$

Define $2n \times 2n$ matrix U as:

$$U = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \tag{3.102}$$

So one have the following:

$$\int_0^l [(k_1''')^2(x) + (k_2''')^2(x)]dx = \mathbf{k}^T U \mathbf{k} \tag{3.103}$$

Finally, the performance index is a linear-quadratic form of $\mathbf{k}$:

$$\begin{aligned} J &= \int_0^\infty \epsilon(t)^2 dt + r\int_0^l [(k_1''')^2(x) + (k_2''')^2(x)]dx \\ &= \mathbf{k}^T R \mathbf{k} - 2P\mathbf{k} + T \end{aligned} \tag{3.104}$$

where

$$R = \gamma^2 HVH^T + GM_1^T VD + GDVM_1 + rU \tag{3.105}$$

$$P = \gamma \mathbf{1}^T VH^T - GM_2^T VD^T - \beta B^T M_1 \tag{3.106}$$

$$T = \mathbf{1}^T V \mathbf{1} + 2\beta B^T M_2 - \frac{\beta^2}{2F}. \tag{3.107}$$

According to standard results on linear-quadratic optimization, if R is nonnegative, $\mathbf{k}_0$ minimizes J if and only if

$$R\mathbf{k}_0 - P^T = 0. \tag{3.108}$$

Since P^T is in range space of R, there exist at least one $\mathbf{k}_0$ that minimize J. If R is positive definite, $\mathbf{k}_0$ is unique. From the numerical result, R is ill-conditioned, so $\mathbf{k}_0$ is not unique. The singular value decomposition method is used and $\mathbf{k}_0$ is chosen to be the minimum norm solution.

3.3 Examples

The control law $y = \mathbf{l}^T\mathbf{x}$ is chosen to minimize

$$J = \int_0^\infty (\, q_0\langle Q\mathbf{x}, \mathbf{x}\rangle + r_0\langle u, u\rangle \,)dt \tag{3.109}$$

where q_0/r_0 is 3000 and

$$Q = \begin{bmatrix} A_{0n} & 0 \\ 0 & I_{n\times n} \end{bmatrix}. \tag{3.110}$$

The plant model and the functional observer are defined in Sections 2.3 and 3.1 respectively. The length of the Euler-Bernoulli beam is 1.

Two sensors are distributed from 0.2 to 0.6. The actuator B_0 is:

$$B_0(x) = \begin{cases} \sin 5\pi(x - 0.3) & \text{if } 0.3 \le x \le 0.5 \\ 0 & \text{otherwise.} \end{cases} \tag{3.111}$$

Figures 3.11-3.14 demonstrate the influence of weighting r on k_1'' and k_2''. Both k_1'' and k_2'' are convergent due to penalizing $r\int_0^l [(k_1''')^2(\eta) + (k_2''')^2(\eta)]d\eta$ in performance index. Larger r can make them smoother but result in large open-loop error, as illustrated in Tables 3.2-3.3.

Let n be the number of modes in the finite dimensional model from which the optimal sensor gains $\mathbf{k}_1$, $\mathbf{k}_2$ are obtained and N be the number of modes of finite dimensional approximation for the plant in application. Figures 3.15-3.17 are the open-loop magnitude response of the observer output and the desired control law, $n = N$ for each case. The difference between the two curves in each of Figures 3.15-3.17 is the square of L_2 norm of error $\|\hat{\epsilon}(j\omega)\|_2^2$ in frequency domain. The value of $\|\hat{\epsilon}(j\omega)\|_2^2$ is equal to $\|\epsilon(t)\|_2^2$ and is shown in Tables 3.2-3.3 for different r and n ($n = N$). The magnitude response is an open-loop response because the observer is only cascaded to the plant

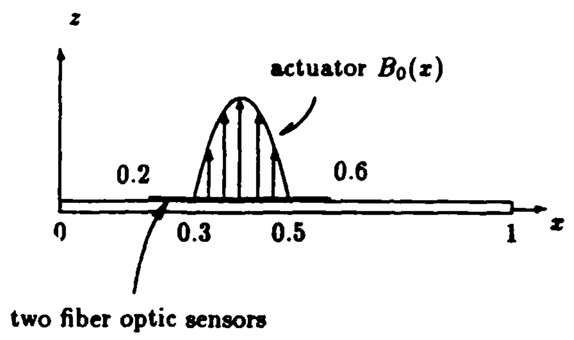

Figure 3.10: The actuator $B_0(x)$ and the distribution of collocated sensors

and the observer output does not feed back to the plant. Therefore it is the open-loop error that we minimize.

Figures 3.18-3.20 are the closed-loop magnitude response, with $n = N$. The input is $e^{i\omega t}$ acting at the actuator function B_0. The output is the transverse displacement at $x = 0.4$. In each plot the solid line denotes the response of the system whose compensator using a functional observer in the feedback loop. The dotted line represents the ideal full state feedback control. Comparing two curves, we know whether the observer is efficient or not.

Figures 3.23 and 3.24 are the closed-loop magnitude responses with $n = 5, N = 15$ and $n = 10, N = 15$, respectively. In Figure 3.23, the first mode is lightly damped as compared to other plots. It is because the real part of closed-loop eigenvalue of observer $n = 5$ is only 19.4% of that of observer $n = 15$, as shown in Table 3.1.

At high frequencies, the observer achieve high performance. This can be verified by using different r's, n's and N's.

	first mode eigenvalue	
	real part	imaginary part
$n = 5, N = 15$	-1.220	8.120
$n = 10, N = 15$	-3.764	6.369
$n = 15, N = 15$	-6.285	6.671
full state feedback	-6.227	9.719

Table 3.1: First mode eigenvalues

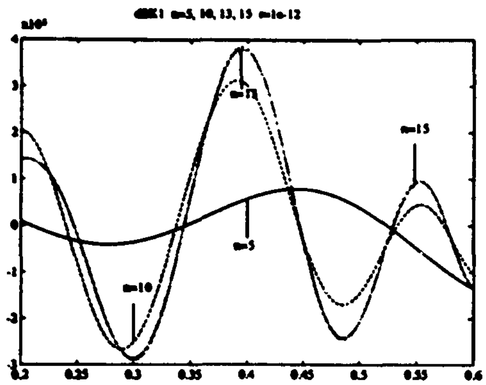

Figure 3.11: k_1'', $r = 10^{-12}$, $n = 5, 10, 13, 15$

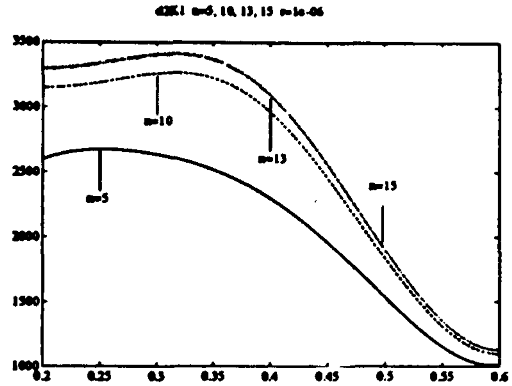

Figure 3.13: k_1'', $r = 10^{-6}$, $n = 5, 10, 13, 15$

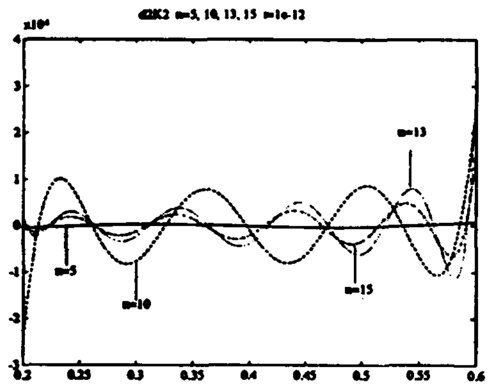

Figure 3.12: k_2'', $r = 10^{-12}$, $n = 5, 10, 13, 15$

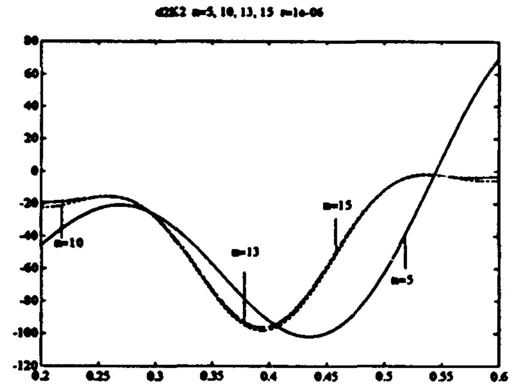

Figure 3.14: k_2'', $r = 10^{-6}$, $n = 5, 10, 13, 15$

	$\|\epsilon\|_2^2$ =ERROR
$n = 5, N = 5$	0.013
$n = 10, N = 10$	38.644
$n = 13, N = 13$	49.147
$n = 15, N = 15$	49.701

Table 3.2: $r = 10^{-12}$

	$\|\epsilon\|_2^2$ =ERROR
$n = 5, N = 5$	0.711×10^3
$n = 10, N = 10$	1.742×10^3
$n = 13, N = 13$	1.963×10^3
$n = 15, N = 15$	1.982×10^3

Table 3.3: $r = 10^{-6}$

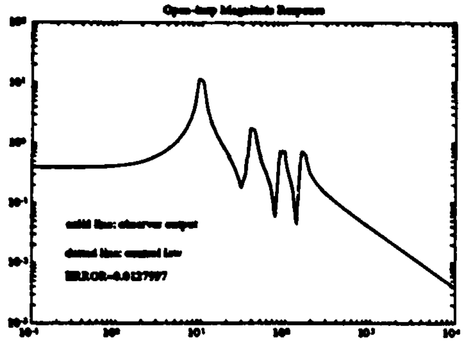

Figure 3.15: Open-loop magnitude responses of observer output φ and control law y, $n = 5$, $N = 5$, $r = 10^{-12}$

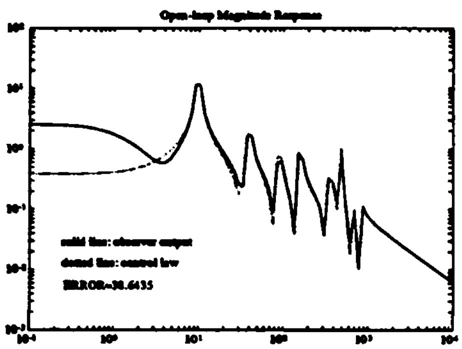

Figure 3.16: Open-loop magnitude responses of observer output φ and control law y, $n = 10$, $N = 10$, $r = 10^{-12}$

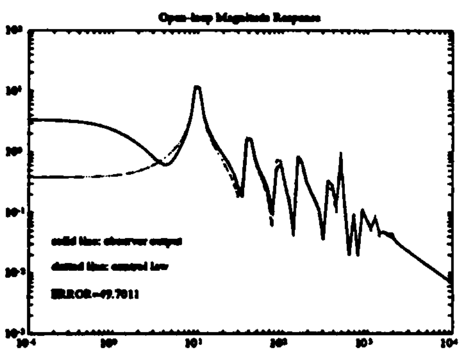

Figure 3.17: Open-loop magnitude responses of observer output φ and control law y, $n = 15$, $N = 15$, $r = 10^{-12}$

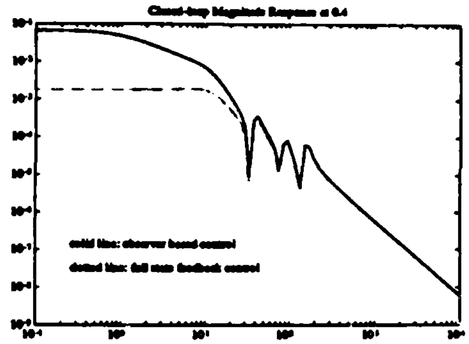

Figure 3.18: Closed-loop magnitude response, $n = 5, N = 5$, $r = 10^{-12}$

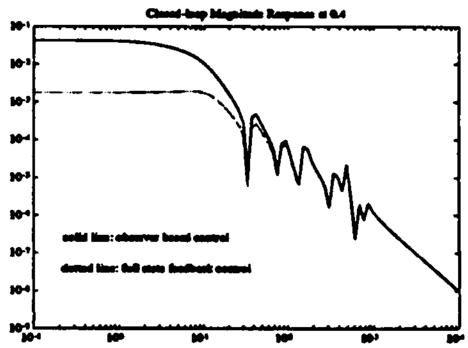

Figure 3.19: Closed-loop magnitude response, $n = 10, N = 10$, $r = 10^{-12}$

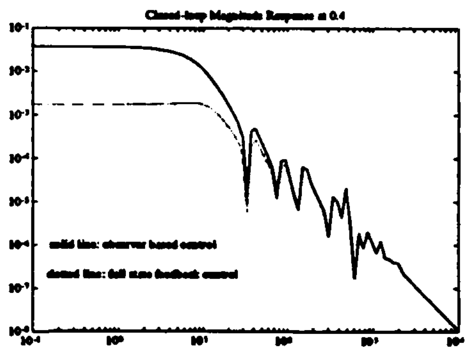

Figure 3.20: Closed-loop magnitude response, $n = 15, N = 15$, $r = 10^{-12}$

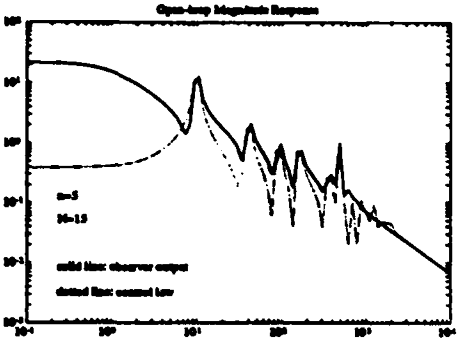

Figure 3.21: Open-loop magnitude responses of observer output φ and control law y, $n = 5$, $N = 15$, $r = 10^{-12}$

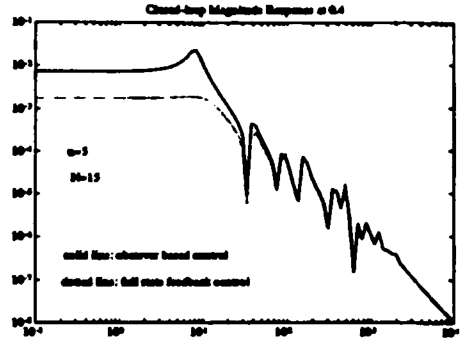

Figure 3.23: Closed-loop magnitude response, $n = 5, N = 15, r = 10^{-12}$

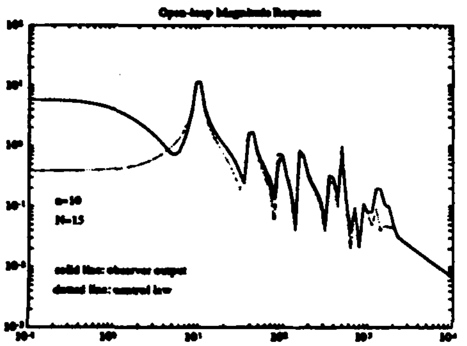

Figure 3.22: Open-loop magnitude responses of observer output φ and control law y, $n = 10$, $N = 15$, $r = 10^{-12}$

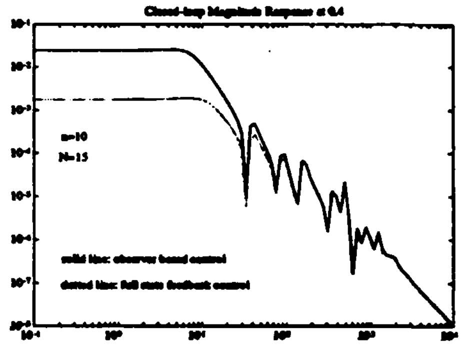

Figure 3.24: Closed-loop magnitude response, $n = 10, N = 15, r = 10^{-12}$

4 Conclusions

A first-order functional observer for an optimal control law for the distributed model of a one-dimensional flexible structure, such as a beam or rod, can be constructed if the outputs of two properly designed fiber optic sensors are available. The sensor designs are determined by the functional gains in the control law determined from control theory for distributed systems. Usually, these sensors must extend over the entire length of the structure. Theorem 1 in Section 2 provides conditions that the sensor gains must satisfy. The most important one is the existence condition. If the existence condition is satisfied, it follows from [6] that the finite dimensional sensor gains will converge as the number of modes in the approximation scheme increases.

When the sensors extend over subintervals of a structure, the sensors are designed to minimize the error between the desired control law that the control law the firs-order observer constructs from the sensor outputs. This design criterion leads to a quadratic optimization problem for any finite number of modes in the model of the structure, and a small penalty (i.e., regularization) on appropriate derivatives of the functions defining how the sensors are tapered guarantees that the sensor designs converge as the number of modes in the model increase. Section 3 formulates the optimization problem and demonstrates the design procedures in detail. The numerical examples demonstrate that the resulting closed-loop control system achieves high performance.

References

[1] M.J. Balas. Feedback control of flexible systems. *IEEE Trans. AC*, AC-23:673–679, 1978.

[2] C.D. Butter and G.B. Hocker. Fiber optics strain gage. *Applied Optics*, 17(18):2867–2869, September 1978.

[3] David E. Cox, Douglas K. Lindner, C. Zackery Furness, and Stanjoe Bingulac. Identification and control of a flexible beam using modal domain sensor. In *Fiber Optic Smart Structures and Skins III*, volume 1370, pages 271–282. SPIE, 1990.

[4] D.E. Cox and D.K. Lindner. Active control for vibration suppression in a flexible beam using a modal domain optical fiber sensor. *ASME Journal of Vibration and Acoustics*, 113:369–382, July 1991.

[5] Thomas G. Giallorenzi, Joseph A. Bucaro, Anthony Dandridge, Jr. G. H. Sigel, James H. Cole, Scott C. Rashleigh, and Richard G. Priest. Optical fiber technology. *IEEE Journal of Quantum Electronics*, QE-18(4):626–664, April 1982.

[6] J.S. Gibson and A. Adamian. Approximation theory for linear-quadratic-gaussian optimal control of flexible structures. *SIAM Journal of Control and Optimization*, 29(1):1–37, January 1991.

[7] J.S. Gibson and A. Adamian. A comparison of three approximation schemes for optimal control of a flexible structure. In H.T. Banks, editor, *Frontiers in Applied Mathematics*, volume 11, chapter 3, pages 85–124. SIAM, Philadelphia, 1992.

[8] J.S. Gibson and C.-L. Meng. Design of fiber optic sensors for control of flexible structures. In *American Control Conference*, volume 2, pages 1686–1688, 1992.

[9] C. Liguori and M. Martinelli. Integral phase modulation properties of a single- mode optical fiber subjected to controlled vibration. *Applied Optics*, 20(24):4319–4323, December 1981.

[10] D.K. Lindner, G.A. Zvonar, W.T. Baumann, and P.L. Delos. Nonlinear effect of a modal domain optical fiber sensor in a vibration suppression control loop for a flexible structure. *ASME Journal of Vibration and Acoustics*, 115:120–128, January 1993.

[11] Douglas K. Lindner, Karl M. Reichard, and William T. Baumann. Measurement and control of flexible structures using distributed sensors. In *Proceedings of the 29th Conference on Decision and Control*, pages 2588–2592. IEEE, December 1990.

[12] Mario Martinelli. The dynamic behavior of a single-mode optical fiber strain gage. *IEEE Journal of Quantum Electronics*, QE-18(4):666–670, April 1982.

[13] Ching-Ling Meng. *Identification and Control of Mechanical and Optical Systems Using Distributed Sensing*. PhD thesis, UCLA, December 1993.

[14] Takanori Okoshi. *Optical Fibers*. Academic Press Inc., 1982.

[15] D.L. Russell. On mathematical models for the elastic beam with frequency-proportional damping. In H.T. Banks, editor, *Frontiers in Applied Mathematics*, volume 11, chapter 4, pages 125–169. SIAM, Philadelphia, 1992.

[16] James S. Sirkis and Henry W. Haslach, Jr. Interferometric strain measurement by arbitrarily configured, surface-mounted, optical fiber. *Journal of Lightwave Technology*, 8(10):1497–1503, October 1990.

[17] W.M. Weber, T.D. Wang, and C.M. Dube. Determination of structural modal amplitude coefficients from a composite-embedded fiber optic sensor using Kalman filter algorithm. In *Fiber Optic Smart Structures and Skins III*, volume 1370, pages 296–305. SPIE, 1990.

Least-Squares and Minimax Parameter Estimation for Linear Systems in the Presence of Noise *

J.S. Gibson and G.H. Lee
Mechanical, Aerospace, and Nuclear Engineering
University of California, Los Angeles 90024-1597

Abstract

This paper addresses the problem of fitting digital input/output models to data generated by linear systems in the presence of white process and sensor noise. The systems of interest have state-space realizations in Hilbert spaces. Both finite-dimensional and infinite-dimensional input/output models are considered. The paper derives a number of new results for least-squares estimation and filtering, and introduces a new class of minimax parameter-estimation and filtering problems. The main results characterize the asymptotic values to which parameter estimates converge with increasing amounts of data.

*This research was supported by AFOSR Grant 910016.

1 Introduction

The linear systems, or plants, considered in this paper have time-invariant discrete-time realizations in possibly infinite-dimensional Hilbert spaces. The plants are driven by unknown white process-noise sequences and possibly by known forcing sequences. Also, the measured outputs are contaminated by white sensor-noise. The objective of the paper is to characterize the asymptotic values of estimated parameters in digital input/output models when such models are fit to the data according to least-squares (LS) estimation criteria and according to new minimax estimation criteria introduced in the paper.

We consider one-step-ahead output prediction, primarily with ARX models, which are models containing linear regressions of measured outputs and known inputs. Also, two results are given on least-squares estimation of FIR and IIR models, which use only histories of known inputs to predict a measured output. The main results of the paper for LS estimation show that, under common conditions, the asymptotic least-squares parameter estimates for ARX models are Markov parameters of Kalman filters that would be constructed for the plant were the state-space model and the noise statistics known. The hypotheses about the noise processes are consistent with those in Kalman filtering theory for finite-dimensional systems and the analogous infinite-dimensional theory [1, 2]. However, constructing a Kalman filter requires the plant and noise statistics to be known, whereas, in the parameter estimation problems considered here, the plant and noise are unknown; only input/output data is available.

Since we are concerned with the limits to which parameter estimates converge as increasing amounts of data are used, we require that the plant be exponentially stable so that the response approaches a steady state, or stationary response. Also, the known forcing sequence is required to have asymptotically time-invariant statistics so that the system response will have asymptotically time-invariant statistics. Most of the results require the known forcing sequence to be white. Such conditions are realistic because it is common in experimental identification to excite a system with a known white sequence or a band-limited sequence obtained by filtering a white sequence; in the latter case, the band-pass filter used in obtaining the band-limited forcing sequence becomes part of the plant to be identified.

An important result that does not require the known forcing sequence to be white concerns the problem of using LS estimation to fit an infinitely-long ARX model to input/output data from a plant of the class considered here. In this case, the asymptotic parameter estimates are the Markov parameters of a steady-state (infinite-time) Kalman filter. This result is known for finite-dimensional plants, and is the basis for the Observer/Kalman Identification (OKID) method developed in [3, 4, 5]. This paper obtains the corresponding result for infinite dimensions, but more important are the results here for finite-dimensional ARX models, since only finite-dimensional models can be used in practice.

It is shown that, under common conditions, using LS estimation to fit an ARX model of order N to input/output data from either a finite-dimensional or an infinite-dimensional plant produces asymptotic parameter estimates that are equal to Markov parameters of the discrete-time Kalman filter that solves a certain minimum-variance state-space filtering problem on an interval of length N. The dimension of this state-space filter equals the dimension (possibly infinite) of the plant generating the data (even though this dimension is not known to the parameter-estimation algorithm), but the length of the time interval for which the state-space filter is optimal equals the order of the ARX model identified. Hence, as the order of the ARX model becomes large, the corresponding Kalman filter approaches the steady-state Kalman filter. (By Markov parameters for a filter with time-dependent gains, we mean the coefficients of past input and output data in a prediction formula, an obvious generalization from the time-invariant case. See Section 5.1.) Another important point is that, while the asymptotic parameter estimates for the infinite-dimensional ARX model are independent of the signal-to-noise ratio of the input/output data, this is not true of the parameter estimates for a finite-dimensional ARX model. The results here show how the signal-to-noise ratio

affects the asymptotic parameter estimates for a finite-dimensional ARX model.

A new class of parameter estimation problems is introduced in Section 4, where the estimated parameters are minimizing solutions to minimax problems for quadratic fit-to-data criteria. The parameters that solve the order-recursive minimax problem in Section 4 are Markov parameters of a discrete-time H_∞ filter for a finite time interval. Our purpose in introducing the minimax parameter estimation problems is to establish the basis for adaptive filters that share the disturbance-attenuation and robustness properties of H_∞ filters designed for known state-space models [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

In Section 4, we consider only finite-dimensional ARX models because numerical parameter-estimation algorithms are possible for the finite-dimensional models only. Under the hypotheses of Section 4, known results on linear-quadratic games imply that the Markov parameters of an H_∞ filter for a finite time interval approach the Markov parameters of a steady-state H_∞ filter as the length of the interval becomes large. Therefore, it follows from the results in this paper that, as the order of the ARX model becomes large, the minimax parameter estimates approximate Markov parameters of a steady-state H_∞ filter.

This paper does not address numerical algorithms for solving the parameter-estimation problems analyzed here. There are, of course, many sophisticated methods for both adaptive and off-line least-squares parameter estimation (adaptive lattice, transversal, and square-root filters, adaptive and nonadaptive QR algorithms). For the minimax parameter-estimation problems introduced in this paper, we have begun to develop adaptive (i.e., recursive) numerical methods, but no methods with the efficiency and stability of modern numerical methods for LS problems exist yet for the minimax problems. Fast, numerically stable algorithms should come soon. Meanwhile, the purpose of this paper with regard to minimax parameter estimation is to define the new problems and characterize the asymptotic properties of the solutions.

The analytical methods of this paper are influenced by the following practical questions, which underlie the entire paper. When we run a parameter-estimation algorithm with a particular set of experimental data, what happens? Do the parameter estimates converge? If so, to what? Although our assumptions about the whiteness and independence of process and sensor-noise sequences are essential to the methods and results, we view the central questions and answers of the paper as fundamentally deterministic. Thus the analytical methods of the paper are purely deterministic. It can be argued that the white-noise sequences that we assume are sample sequences from ergodic stochastic processes, but the central questions of this paper—about what happens with a particular experimental data sequence—can be answered in a deterministic framework. We believe that, by stripping away the probabilistic machinery usually employed when studying problems with random noise (see [19] for example), we cut more directly to the questions addressed here about what happens with actual data.

We set up the basic analytical framework in Section 2 by defining correlation, independence, and whiteness for scalar and Hilbert-space-valued sequences. These definitions, which involve time averages, are standard for scalar sequences, and the generalizations to Hilbert-space-valued sequences are straightforward. The novel component of this framework is a Hilbert space containing equivalence classes of deterministic scalar sequences. The parameter-estimation problems are formulated and analyzed in this space. Because we want to use the correlation of deterministic scalar sequences as an inner product, the Hilbert space must contain equivalence classes instead of scalar sequences themselves. Two scalar sequences are defined to be equivalent if their difference is asymptotically zero in the mean-square sense. The correlation of two scalar sequences is the inner product of their respective equivalence classes. In particular, uncorrelated sequences (i.e., sequences with correlation 0) belong to orthogonal equivalence classes.

We define and analyze parameter-estimation problems for infinite data sequences in Sections 3 (least-squares problems) and Section 4 (minimax problems). While the most important interpre-

tation of the paper's main results is the characterization of the solutions to certain parameter estimation problems for ARX models of an unknown plant in terms of Kalman and H_∞ filters for state-space models of a known plant, we do not employ state-space filtering theory, stochastic or otherwise, in our basic analysis. Rather, the framework of white deterministic sequences established in Section 2 makes it easy to show that the parameter estimation problems are equivalent to certain linear-quadratic (LQ) control problems. The LS estimation problems are equivalent to linear-quadratic regulator (LQR) problems, and the minimax estimation problems are equivalent to linear-quadratic games.

Section 5 shows the precise relationships of the estimated ARX parameters to state-space filters. This is the only section of the paper that uses any state-space filtering theory. It is most convenient to characterize the solutions to the parameter estimation problems first in terms of the detailed solutions to the corresponding optimal control problems: Riccati operators, closed-loop systems, etc. (These details are not discussed in Sections 3 and 4.) The relationships of the estimated parameters to Kalman and H_∞ filters follow then from the dual mathematical structures of optimal linear-quadratic controllers and optimal state-space filters.

Section 6 discusses parameter-estimation problems with finite data sequences, the only parameter-estimation problems that can be solved numerically. These problems are just restatements of the problems in Sections 3 and 4 except that the objective functionals in Section 6 are defined for finite data sequences. The main motivation for the framework set up in Section 2 is to allow the definitions in Sections 3 and 4 of parameter-estimation problems to which the problems in Section 6 converge as the lengths of the finite data sequences increase. According to the theorems in Section 6, the parameters that solve the problems with finite data sequences converge to the parameters that solve the corresponding problems with infinite data sequences. This convergence follows easily from the problem definitions in Sections 3 and 4.

2 Statistics of Deterministic Sequences

All sequences in this paper will have the form

$$y = [y_0 \; y_1 \; y_2 \; \ldots]. \tag{2.1}$$

The shift operator q^{-1} is defined by

$$q^{-1}y = [0 \; y_0 \; y_1 \; y_2 \; \ldots], \tag{2.2}$$

and q^{-n} means $(q^{-1})^n$ for any nonnegative integer n.

For y and v scalar sequences, we define

$$\langle\!\langle y, v \rangle\!\rangle = \lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t} y_i \bar{v}_i, \tag{2.3}$$

$$\|y\| = \langle\!\langle y, y \rangle\!\rangle^{1/2} = [\lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t} |y_i|^2]^{1/2} \tag{2.4}$$

when the limits exist. (We say that such limits exist only if they are finite.) Two scalar sequences y and v have *correlation* α if the limit in (2.3) exists and equals the (finite) complex number α; y and v are *independent* if $q^{-m}y$ and $q^{-n}v$ have correlation 0 for all nonnegative integers m and n.

Let w and u be sequences in Hilbert spaces W and U, respectively. For each $\eta \in W$,

$$[\![w, \eta]\!] = [\langle w_0, \eta \rangle_W \;\; \langle w_1, \eta \rangle_W \;\; \langle w_2, \eta \rangle_W \;\; \ldots] \tag{2.5}$$

is a scalar sequence. The sequences w and u have *bounded correlation* if, for each $\eta \in W$ and $\xi \in U$, the scalar sequences $[\![w, \eta]\!]$ and $[\![u, \xi]\!]$ have some correlation α. When this is the case, the Principle of Uniform Boundedness implies that there exists a real number r such that

$$|\langle\!\langle [\![u, \xi]\!], [\![w, \eta]\!] \rangle\!\rangle| \leq r^2 |\xi| \cdot |\eta| \qquad \forall \eta \in W,\ \xi \in U. \tag{2.6}$$

When w and u have bounded correlation, there exists a unique bounded linear operator R^{uw} from W to U such that

$$\langle\!\langle [\![u, \xi]\!], [\![w, \eta]\!] \rangle\!\rangle = \langle R^{uw}\eta, \xi \rangle_U. \tag{2.7}$$

The operator R^{uw} is called the *correlation* operator for u and w. The Hilbert-space-valued sequence w has *bounded autocorrelation* if the conditions for w and u to have bounded correlation hold with $u = w$. In this case, we write R^w for R^{ww}. The operator R^w is called the *autocorrelation* operator for w, and R^w is self-adjoint and nonnegative. (Operators analogous to the correlation operators here have been defined for Hilbert-space-valued random variables. See [1, 2] for example.)

The Hilbert-space-valued sequences w and u are *independent* if the scalar sequences $[\![w, \eta]\!]$ and $[\![u, \xi]\!]$ are independent for all $\eta \in W$ and $\xi \in U$. Hence w and u are independent if and only if $q^{-m}w$ and $q^{-n}u$ have bounded correlation and zero correlation operator for all nonnegative integers m and n.

A Hilbert-space-valued sequence w is *white* if w has bounded autocorrelation and w and $q^{-n}w$ have bounded correlation and zero correlation operator for all positive integers n.

In the following sections, the input sequences to linear systems will satisfy at least the following condition for a sequence u in a Hilbert space U:

$$q^{-i}u \text{ and } q^{-j}u \text{ have bounded correlation for all nonnegative integers } i \text{ and } j. \tag{2.8}$$

Theorem 2.1 (Convolution) *Suppose that u satisfies* (2.8), $R^u \le r^2$ *for some* $r \ge 0$, *and* $a_i \in U$ *with*

$$\sum_{i=0}^{\infty} |a_i| < \infty. \tag{2.9}$$

Define the scalar sequence y by

$$y_k = \sum_{i=0}^{\infty} \langle u_{k-i}, a_i \rangle, \qquad k = 0, 1, \ \ldots, \tag{2.10}$$

where $u_{k-i} = 0$ *when* $i > k$. *For this y, the limit in* (2.4) *exists and the following inequalities hold:*

$$\| y \| \le r \sum_{i=0}^{\infty} |a_i|, \tag{2.11}$$

$$\| y - \sum_{i=0}^{n} [\![q^{-i}u, a_i]\!] \| \le r \sum_{i=n+1}^{\infty} |a_i|, \qquad n = 0, 1, \ \ldots \ . \tag{2.12}$$

If v is a scalar sequence with bounded autocorrelation and if v and $q^{-i}u$ *have bounded correlation for all nonnegative integers i, then the limit in* (2.3) *exists.*

Proof Define

$$\langle\!\langle y, v \rangle\!\rangle_t = \frac{1}{t} \sum_{i=0}^{t} y_i \bar{v}_i, \qquad \| y \|_t = [\frac{1}{t} \sum_{i=0}^{t} |y_i|^2]^{1/2}, \qquad t = 1, 2, 3, \ \ldots, \tag{2.13}$$

$$y^n = \sum_{i=0}^{n} [\![q^{-i}u, a_i]\!], \qquad n = 0, 1, 2, \ \ldots \ . \tag{2.14}$$

The Principle of Uniform Boundedness and (2.8) imply that there exists a real number ρ such that

$$\| [\![q^{-i}u, \xi]\!] \|_t \le \rho |\xi| \qquad \forall \xi \in U, \ i \ge 0, \ t \ge 1, \tag{2.15}$$

$$\| y^n - y^m \|_t \le \rho \sum_{i=m+1}^{n} |a_i|, \qquad \forall n \ge m, \ t \ge 1. \tag{2.16}$$

Also, $R^u \le r^2$ implies

$$\| [\![q^{-i}u, \xi]\!] \| \le r |\xi| \qquad \forall \ i \ge 0, \tag{2.17}$$

$$\| y^n - y^m \| \le r \sum_{i=m+1}^{n} |a_i|, \qquad \forall n \ge m, \tag{2.18}$$

Hence, $\| y^n \|$ converges to some real μ. For $m, n \ge 0$,

$$\big| \| y^n \|_t - \| y^n \| \big| \le \| y^n - y^m \|_t + \big| \| y^m \|_t - \| y^m \| \big| + \| y^m - y^n \|. \tag{2.19}$$

Since $\lim_{t \to \infty} \| y^m \|_t = \| y^m \|$ for each m, (2.16)—(2.19) imply that, as $t \to \infty$, $\| y^n \|_t \to \| y^n \|$ uniformly in n. That the limit in (2.4) exists follows then from

$$\mu = \lim_n \lim_t \| y^n \|_t = \lim_t \lim_n \| y^n \|_t = \lim_t \| y \|_t = \| y \|. \tag{2.20}$$

Since $\lim_t \| y^n \|_t = \| y^n \|$, (2.14), (2.17), and (2.20) imply (2.11). Then replacing y in (2.11) with $y - y^n$ yields (2.12).

Since v and $q^{-i}u$ have bounded correlation for all nonnegative i, v and y^m have bounded correlation for all m, and (2.18) implies that $\langle\!\langle y^m, v\rangle\!\rangle$ is a Cauchy sequence. For $n > t$, $\langle\!\langle y, v\rangle\!\rangle_t = \langle\!\langle y^n, v\rangle\!\rangle_t$. Let $\epsilon > 0$, and choose m such that the right-hand side of (2.16) is less than ϵ. Choose t such that $|\langle\!\langle y^m, v\rangle\!\rangle_t - \langle\!\langle y^m, v\rangle\!\rangle| < \epsilon$, and let $n = t + m$. Then

$$|\langle\!\langle y, v\rangle\!\rangle_t - \langle\!\langle y^m, v\rangle\!\rangle| \leq |\langle\!\langle y^n - y^m, v\rangle\!\rangle_t| + |\langle\!\langle y^m, v\rangle\!\rangle_t - \langle\!\langle y^m, v\rangle\!\rangle| \leq (\sup_t \|v\|_t + 1)\epsilon. \tag{2.21}$$

Therefore,

$$\lim_t \langle\!\langle y, v\rangle\!\rangle_t = \lim_m \langle\!\langle y^m, v\rangle\!\rangle. \qquad \square \tag{2.22}$$

When u satisfies (2.8), we define S_u to be the set of all scalar sequences y having the form in (2.10) with a_i satisfying (2.9). Then S_u is a linear space, and Theorem 2.1 implies that (2.3) and (2.4) define, respectively, a sesquilinear functional and a seminorm on S_u.

Definition 2.2 *Two scalar sequences y and v are* equivalent *if $\|y - v\| = 0$. Two sequences w and $\tilde{w}$ in a Hilbert space W are* equivalent *if $[\![w, \eta]\!]$ and $[\![\tilde{w}, \eta]\!]$ are equivalent $\forall \eta \in W$.*

Theorem 2.3 *If w and $\tilde{w}$ are equivalent sequences in a Hilbert space W and u and $\tilde{u}$ are equivalent sequences in a Hilbert space U, then w and u have bounded correlation and correlation operator R^{uw} if and only if $\tilde{w}$ and $\tilde{u}$ have bounded correlation and correlation operator R^{uw}.*

When u satisfies (2.8), we define $\mathcal{S}_u$ to be the set of all equivalence classes of elements of S_u, so that $\mathcal{S}_u$ is an inner-product space, or pre-Hilbert space, with inner product and norm determined by (2.3) and (2.4), respectively. If u and w are sequences in Hilbert spaces U and W, respectively, if u and w each satisfy (2.8), and if $q^{-i}u$ and $q^{-j}w$ have bounded correlation for all nonnegative integers i and j, then $\mathcal{S}_u \oplus \mathcal{S}_w$ is an inner-product space with inner product and norm determined by (2.3) and (2.4).

When u satisfies (2.8), we define the Hilbert space $\tilde{\mathcal{S}}_u$ to be the completion of $\mathcal{S}_u$. Hence the Hilbert space $\tilde{\mathcal{S}}_u \oplus \tilde{\mathcal{S}}_w$ is the completion of $\mathcal{S}_u \oplus \mathcal{S}_w$ (under the conditions stated for $\mathcal{S}_u \oplus \mathcal{S}_w$ to be an inner product space). If u and w are independent, $\mathcal{S}_u$ and $\mathcal{S}_w$ are orthogonal. Whether every element of $\tilde{\mathcal{S}}_u$ is an equivalence class of scalar sequences is a open question, which does not matter for the rest of this paper because the only elements of $\tilde{\mathcal{S}}_u \oplus \tilde{\mathcal{S}}_w$ encountered are obtained from convolutions of the generating sequences u and w.

Henceforth, the notation in this paper usually will not distinguish between a scalar sequence s and the equivalence class of scalar sequences equivalent to s when s is a data or noise sequence. This abuse of notation is similar to the common practice of referring to elements of the Hilbert space $L_2(0,1)$ as square-integrable functions. On occasion, we will refer to the following condition for a Hilbert-space-valued sequence u:

$$[\![q^{-i}u, \xi]\!] \notin \text{closed span}\{[\![q^{-j}u, \hat{\xi}]\!] : j \geq 0,\, j \neq i\} \qquad \forall i \geq 0 \text{ and } \forall \xi, \hat{\xi} \in U, \tag{2.23}$$

where $u_k \in U$ ($k = 0, 1, 2, \ldots$) and closed span$\{[\![q^{-j}u, \hat{\xi}]\!] : j \geq 0,\, j \neq i\}$ is a subset of $\tilde{\mathcal{S}}_u$. While the technically correct interpretation of (2.23) requires that $[\![q^{-i}u, \xi]\!]$ and $[\![q^{-j}u, \hat{\xi}]\!]$ be interpreted as equivalence classes of scalar sequences, the practical meaning of (2.23) is that, for any two elements ξ and $\hat{\xi}$ of U, the scalar sequence $[\![q^{-i}u, \xi]\!]$ cannot be approximated arbitrarily closely in the sense of the seminorm in (2.4) by linear combinations of the scalar sequences $[\![q^{-j}u, \hat{\xi}]\!]$, $j \neq i$. The condition (2.23) holds if, for example, u is a finite linear combination of periodic sequences and a nonzero white sequence.

In the rest of this paper, $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ and $\|\cdot\|$ will denote the inner product and norm on $\tilde{\mathcal{S}}_u \oplus \tilde{\mathcal{S}}_w$, while $\langle \cdot, \cdot \rangle$ and $|\cdot|$, without subscripts, will denote the inner products and norms, respectively, on all other Hilbert spaces (including C, the complex plane). There should be no ambiguity about which elements belong to which spaces.

3 Least-Squares Estimation and Prediction

We let U, W, and Y be Hilbert spaces, and we assume that there exist absolutely summable sequences of bounded linear operators $L_k^{yu} \in \mathcal{B}(U,Y)$, $L_k^{yw} \in \mathcal{B}(W,Y)$ and absolutely summable sequences $c_k^{su} \in U$, $c_k^{sw} \in W$, $a_k^{su} \in U$, and $a_k^{sy} \in Y$. For each k, we define $A_k^{su} \in U'$, $A_k^{sy} \in Y'$, $L_k^{su} \in U'$, and $L_k^{sw} \in W'$ by

$$A_k^{su}\eta = \langle \eta, a_k^{su} \rangle, \qquad A_k^{sy}\eta = \langle \eta, a_k^{sy} \rangle, \qquad L_k^{su}\eta = \langle \eta, c_k^{su} \rangle, \qquad L_k^{sw}\eta = \langle \eta, c_k^{sw} \rangle, \tag{3.1}$$

where, in each case, η is in U, W, or Y as needed. (When U, for example, is $\mathbf{C}^n$, each a_k^{su} is an n-dimensional column vector and A_k^{su} is the complex-conjugate of the transpose of a_k^{su}.)

For u a sequence in U, we write

$$L_k^{yu}u = [L_k^{yu}u_0 \ L_k^{yu}u_1 \ L_k^{yu}u_2 \ \dots], \qquad A_k^{su}u = [A_k^{su}u_0 \ A_k^{su}u_1 \ A_k^{su}u_2 \ \dots], \tag{3.2}$$

and similar equations for L_k^{yw}, L_k^{su}, etc. We define the convolution operators L^{yu}, A^{su}, L^{yw}, L^{sw}, L^{su}, and A^{sy} by

$$L^{yu}u = \sum_{k=0}^{\infty} L_k^{yu}q^{-k}u, \qquad A^{su}u = \sum_{k=0}^{\infty} A_k^{su}q^{-k}u, \qquad \text{etc.} \tag{3.3}$$

Each of these operators maps sequences to sequences. Changing the order of summation in $A^{sy}L^{yu}u$ and $A^{sy}L^{yw}w$ yields

$$A^{sy}L^{yu}u = \sum_{i=0}^{\infty}[\sum_{k=0}^{i} A_k^{sy}L_{i-k}^{yu}]q^{-i}u = \sum_{i=0}^{\infty}[\![q^{-i}u, \sum_{k=0}^{i} L_{i-k}^{yu*}a_k^{sy}]\!], \tag{3.4}$$

$$A^{sy}L^{yw}w = \sum_{i=0}^{\infty}[\sum_{k=0}^{i} A_k^{sy}L_{i-k}^{yw}]q^{-i}w = \sum_{i=0}^{\infty}[\![q^{-i}w, \sum_{k=0}^{i} L_{i-k}^{yw*}a_k^{sy}]\!], \tag{3.5}$$

where $[\![\cdot,\cdot]\!]$ means a scalar sequence defined as in (2.5), and L_i^{yu*} and L_i^{yw*} are the adjoints of L_i^{yu} and L_i^{yw}, respectively. Hence,

$$L^{su}u - A^{su}u - A^{sy}L^{yu}u = \sum_{i=0}^{\infty}[\![q^{-i}u, [c_i^{su} - a_i^{su} - \sum_{k=0}^{i} L_{i-k}^{yu*}a_k^{sy}]]\!], \tag{3.6}$$

$$L^{su}w - A^{sy}L^{yw}w = \sum_{i=0}^{\infty}[\![q^{-i}w, [c_i^{sw} - \sum_{k=0}^{i} L_{i-k}^{yw*}a_k^{sy}]]\!]. \tag{3.7}$$

The following hypothesis is implicit in all subsequent problems and theorems.

Hypothesis 3.1 *There exist independent sequences u, w, and v in the Hilbert spaces U, W, and $\mathbf{C}^n$, respectively. Each of these sequences has bounded autocorrelation, w is white, and u satisfies* (2.8). *There exist a Hilbert space Y and absolutely summable sequences $L_k^{yu} \in \mathcal{B}(U,Y)$, $L_k^{yw} \in \mathcal{B}(W,Y)$, $c_k^{su} \in U$, $c_k^{sw} \in W$, $L_k^{su} \in U'$, and $L_k^{sw} \in W'$ with c_k^{su} related to L_k^{su} and c_k^{sw} related to L_k^{sw} as in* (3.1). *The sequences $L_0^{sw}w$ and $L_0^{yw}w$ are independent of $L_k^{sw}w$ and $L_k^{yw}w$ $\forall k \geq 1$.*

In the subsequent estimation and prediction problems, u is a known input sequence to a linear system, w is an unknown noise sequence, and v is another unknown sequence, often to be estimated. Process noise and measurement noise are components of w. The measurement-noise sequences are $L_0^{sw}w$ and $L_0^{yw}w$. The inner-product space $\mathcal{S}_u \oplus \mathcal{S}_w$ of equivalence classes of scalar sequences is constructed as in Section 2.

From here on, the notation in this paper does not distinguish between a scalar sequence $[\![q^{-i}w, \eta]\!]$ for $\eta \in W$ and the equivalence class of all scalar sequences equivalent to $[\![q^{-i}w, \eta]\!]$. The notation is abused similarly for u and the output sequences y and s in the following problems. Under Hypothesis 3.1, $\xi_i = c_i^{su} - a_i^{su} - \sum_{k=0}^{i} L_{i-k}^{yu*} a_k^{sy}$ and $\eta_i = c_i^{sw} - \sum_{k=0}^{i} L_{i-k}^{yw*} a_k^{sy}$ are absolutely summable sequences in U and W, respectively. Therefore, the right-hand-sides of (3.6) and (3.7) are elements of S_u and S_w, respectively.

Problem 3.2 (LS: Infinite Length ARX) *Let the Y-valued sequence y and the scalar-valued sequence s satisfy*

$$y = L^{yu}u + L^{yw}w, \tag{3.8}$$

$$s = L^{su}u + L^{sw}w + v. \tag{3.9}$$

Find absolutely summable sequences $a^{sy} = [a_0^{sy}\ a_1^{sy}\ a_2^{sy}\ \ldots]$ in Y and $a^{su} = [a_0^{su}\ a_1^{su}\ a_2^{su}\ \ldots]$ in U to minimize

$$J_E(a^{sy}, a^{su}) = \|A^{sy}y + A^{su}u - s\|^2 \tag{3.10}$$

subject to

$$a_0^{sy} = 0. \tag{3.11}$$

Problem 3.3 (LS: Finite Length ARX) *Assume the hypotheses in* Problem 3.2. *Let N be a positive integer, and find sequences a^{sy} in Y and a^{su} in U to minimize $J_{EN}(a^{sy}, a^{su}) = J_E(a^{sy}, a^{su})$ subject to* (3.11) *and*

$$a_k^{sy} = 0, \qquad a_k^{su} = 0, \qquad \forall\, k > N. \tag{3.12}$$

(The functional $J_{EN}(a^{sy}, a^{su})$ is just $J_E(a^{sy}, a^{su})$ defined for parameters a^{sy} and a^{su} that satisfy (3.11) and (3.12).)

Problem 3.4 (LS: IIR) *Assume the hypotheses in* Problem 3.2. *Find an absolutely summable sequence a^{su} in U to minimize*

$$J_E(0, a^{su}) = \|A^{su}u - s\|^2. \tag{3.13}$$

Problem 3.5 (LS: FIR) *Assume the hypotheses in* Problem 3.2. *Let N be a positive integer, and find an absolutely summable sequence a^{su} in U to minimize $J_{EN}(0, a^{su}) = J_E(0, a^{su})$ subject to*

$$a_k^{su} = 0, \qquad \forall\, k > N. \tag{3.14}$$

In Problems 3.2–3.5, the parameter sequences a^{sy} and a^{su} are chosen to minimize the norm (i.e., $\|\cdot\|$) of the prediction-error sequence

$$\varepsilon \;=\; \sum_{i=1}^{N} A_i^{sy} q^{-i} y + \sum_{i=0}^{N} A_i^{su} q^{-i} u \;-\; s \;=\; \sum_{i=1}^{N} [\![q^{-i}y, a_i^{sy}]\!] + \sum_{i=0}^{N} [\![q^{-i}u, a_i^{su}]\!] \;-\; s. \tag{3.15}$$

(Recall (2.5) and (3.1)–(3.3)). The k^{th} term in this sequence is

$$\varepsilon_k \;=\; \sum_{i=1}^{N} \langle y_{k-i}, a_i^{sy}\rangle + \sum_{i=0}^{N} \langle u_{k-i}, a_i^{su}\rangle - \; s_k, \qquad k \geq 0, \tag{3.16}$$

with $y_i = 0$ and $u_i = 0$ for $i < 0$ (recall (2.2)). In Problems 3.2 and 3.4, $N = \infty$; in Problems 3.4 and 3.5, $a^{sy} = 0$.

Problem 3.6 (LQR: Infinite Interval) *Let X be a Hilbert space and $T \in \mathcal{B}(X,X)$, $B^u \in \mathcal{B}(U,X)$, $B^w \in \mathcal{B}(W,X)$, $C^y \in \mathcal{B}(X,Y)$, $C^s \in X'$. Let the spectral radius of T be less than 1, and let c^{sx} be the element of X that satisfies*

$$C^s x = \langle x, c^{sx} \rangle \qquad \forall x. \tag{3.17}$$

(The element c^{sx} is independent of x.) Find a square-summable sequence ν in Y, with $\nu_0 = 0$, such that ν minimizes

$$J_C(\nu) = \sum_{i=1}^{\infty} \langle R^w B^{w*} \xi_i, B^{w*} \xi_i \rangle + \sum_{i=1}^{\infty} \langle L_0^{yw} R^w L_0^{yw*} \nu_i, \nu_i \rangle \tag{3.18}$$

where the sequence ξ in X satisfies

$$\xi_1 = -c^{sx}, \qquad \xi_{k+1} = T^* \xi_k + C^{y*} \nu_k \quad \forall k \geq 1. \tag{3.19}$$

Problem 3.7 (LQR: Finite Interval) *Assume the hypotheses in* Problem 3.6. *Find a sequence ν in Y, with*

$$\nu_0 = 0, \qquad \nu_k = 0 \quad \forall k > N, \tag{3.20}$$

such that ν minimizes

$$J_{CN}(\nu) = \sum_{i=1}^{N} \langle R^w B^{w*} \xi_i, B^{w*} \xi_i \rangle + \sum_{i=1}^{N} \langle L_0^{yw} R^w L_0^{yw*} \nu_i, \nu_i \rangle + \langle G \xi_{N+1}, \xi_{N+1} \rangle \tag{3.21}$$

where the sequence ξ in X satisfies (3.19) *and*

$$G = \sum_{i=0}^{\infty} T^i [B^w R^w B^{w*} + B^u R^u B^{u*}] (T^*)^i. \tag{3.22}$$

The operator G in (3.22) is the unique element of $\mathcal{B}(X,X)$ that satisfies the Lyapunov equation

$$TGT^* - G = -[B^w R^w B^{w*} + B^u R^u B^{u*}]. \tag{3.23}$$

Theorem 3.8 *Assume the hypotheses in* Problems 3.2 *and* 3.6. *Assume further that, for $k \geq 1$,*

$$L_k^{yu} = C^y T^{k-1} B^u, \qquad L_k^{yw} = C^y T^{k-1} B^w, \tag{3.24}$$

$$L_k^{su} = C^s T^{k-1} B^u, \qquad L_k^{sw} = C^s T^{k-1} B^w. \tag{3.25}$$

If sequences a^{sy} and a^{su} solve Problem 3.2, *then the sequence $\nu = a^{sy}$ solves* Problem 3.6. *Conversely, if a sequence ν solves* Problem 3.6, *then the sequence $a^{sy} = \nu$ and the sequence*

$$a_i^{su} = -B^{u*} \xi_i - L_0^{yu*} \nu_i, \qquad i \geq 1, \qquad a_0^{su} = c_0^{su}. \tag{3.26}$$

solve Problem 3.2. *(The sequence ξ is generated by* (3.19).*) If the autocorrelation operator $L_0^{yw} R^w L_0^{yw*}$ is coercive, there exists exactly one sequence ν that solves* Problem 3.6. *If u satisfies* (2.23), *then, for any pair of sequences a^{sy} and a^{su} that solve* Problem 3.2, (3.26) *holds with ξ generated by* (3.19) *with $\nu = a^{sy}$.*

Theorem 3.9 *Assume the hypotheses in* Problems 3.3 *and* 3.7, *assume that* (3.24) *and* (3.25) *hold for $k \geq 1$, and let u be white. If sequences a^{sy} and a^{su} solve* Problem 3.3, *then the sequence $\nu = a^{sy}$ solves* Problem 3.7. *Conversely, if a sequence ν solves* Problem 3.7, *then the sequence $a^{sy} = \nu$ and the corresponding sequence a^{su} given by* (3.26) *solve* Problem 3.3. *If $L_0^{yw} R^w L_0^{yw*}$ is coercive, there exists exactly one sequence ν that solves* Problem 3.6. *If R^u is coercive, then, for any pair of sequences a^{sy} and a^{su} that solve* Problem 3.3, (3.26) *holds with ξ generated by* (3.19) *with $\nu = a^{sy}$.*

Theorem 3.10 *Assume that* (3.24) *and* (3.25) *hold for* $k \geq 1$. *If the hypotheses in* Problems 3.2 *and* 3.6 *hold, then*

$$\min_{\{\nu:\nu_0=0\}} J_E(\nu, a^{su}) = \min_{\nu} J_C(\nu) + \langle R^w c_0^{sw}, c_0^{sw}\rangle + \|v\|^2. \tag{3.27}$$

If the hypotheses in Problems 3.3 *and* 3.7 *hold and u is white, then*

$$\min_{\{\nu:\nu_0=0\}} J_{EN}(\nu, a^{su}) = \min_{\nu} J_{CN}(\nu) + \langle R^w c_0^{sw}, c_0^{sw}\rangle + \|v\|^2. \tag{3.28}$$

Theorem 3.11 *A solution to* Problem 3.4 *is* $a^{su} = c^{su}$. *If u satisfies* (2.23), *the solution to* Problem 3.4 *is unique. If u is white, a solution to* Problem 3.5 *is*

$$a_k^{su} = c_k^{su}, \qquad 0 \leq k \leq N. \tag{3.29}$$

If u is white with R^u *coercive, the solution to* Problem 3.5 *is unique.*

Proof of Theorem 3.8 Under the hypotheses in Problem 3.2,

$$J_E(a^{sy}, a^{su}) = \|A^{sy}L^{yu}u + A^{su}u - L^{su}u\|^2 + \|A^{sy}L^{yw}w - L^{sw}w\|^2 + \|v\|^2, \tag{3.30}$$

$$\|A^{sy}L^{yu}u + A^{su}u - L^{su}u\|^2 = \|\sum_{i=0}^{\infty}[\![q^{-i}u, [c_i^{su} - a_i^{su} - \sum_{k=0}^{i} L_{i-k}^{yu*}a_k^{sy}]]\!]\|^2, \tag{3.31}$$

$$\begin{aligned}\|A^{sy}L^{yw}w - L^{sw}w\|^2 = &\sum_{i=1}^{\infty}\langle R^w[c_i^{sw} - \sum_{k=0}^{i-1} L_{i-k}^{yw*}a_k^{sy}], [c_i^{sw} - \sum_{k=0}^{i-1} L_{i-k}^{yw*}a_k^{sy}]\rangle \\ &+ \sum_{i=0}^{\infty}\langle R^w L_0^{yw*}a_i^{sy}, L_0^{yw*}a_i^{sy}\rangle + \|L_0^{sw}w\|^2.\end{aligned} \tag{3.32}$$

Since c_k^{su}, c_k^{sw}, L_k^{su}, and L_k^{sw} are related as in (3.1), (3.25) implies

$$c_k^{su} = B^{u*}(T^*)^{k-1}c^{sx}, \qquad c_k^{sw} = B^{w*}(T^*)^{k-1}c^{sx}, \qquad k \geq 1. \tag{3.33}$$

Since

$$\xi_i = -(T^*)^{i-1}c^{sx} + \sum_{k=1}^{i-1}(T^*)^{i-1-k}C^{y*}\nu_k, \qquad i > 1, \tag{3.34}$$

it follows from (3.24) that, if $\nu_0 = 0$,

$$c_i^{su} - \sum_{k=0}^{i} L_{i-k}^{yu*}\nu_k = -B^{u*}\xi_i - L_0^{yu*}\nu_i, \qquad i \geq 1, \tag{3.35}$$

$$c_i^{sw} - \sum_{k=0}^{i-1} L_{i-k}^{yw*}\nu_k = -B^{w*}\xi_i, \qquad i \geq 1. \tag{3.36}$$

Hence, if $\nu_0 = 0$,

$$\begin{aligned}J_E(\nu, a^{su}) = &J_C(\nu) + \langle R^w c_0^{sw}, c_0^{sw}\rangle + \|v\|^2 \\ &+ \|[\![u, c_0^{su} - a_0^{su}]\!] + \sum_{i=1}^{\infty}[\![q^{-i}u, (B^{u*}\xi_i + L_0^{yu*}\nu_i + a_i^{su})]\!]\|^2.\end{aligned} \tag{3.37}$$

That $L_0^{yw} R^w L_0^{yw*}$ being coercive is sufficient for the existence of exactly one solution to Problem 3.6 is well known. All of the other statements in the theorem follow from (3.37). $\square$

If u is white also, then

$$\|A^{sy}L^{yu}u + A^{su}u - L^{su}u\|^2 =$$

$$\|[u, c_0^{su} - a_0^{su}] + \sum_{i=1}^{\infty}[q^{-i}u, (B^{u*}\xi_i + L_0^{yu*}a_i^{sy} + a_i^{su})]\|^2 = \tag{3.38}$$

$$\langle R^u(c_0^{su} - a_0^{su}), (c_0^{su} - a_0^{su})\rangle + \sum_{i=1}^{\infty}\langle R^u(B^{u*}\xi_i + L_0^{yu*}\nu_i + a_i^{su}), (B^{u*}\xi_i + L_0^{yu*}\nu_i + a_i^{su})\rangle$$

where ξ satisfies (3.19) with $\nu = a^{sy}$.

Proof of Theorem 3.9 When $\nu_k = 0\ \forall k > N$, (3.19) yields

$$\xi_i = T^{*(i-N-1)}\xi_{N+1} \qquad i \geq N+1. \tag{3.39}$$

Then, when (3.11) holds, (3.37) with $\nu = a^{sy}$ and (3.38) yield

$$\begin{aligned} J_{EN}(a^{sy}, a^{su}) &= \sum_{i=1}^{N}\langle R^w B^{w*}\xi_i, B^{w*}\xi_i\rangle + \sum_{i=1}^{N}\langle L_0^{yw} R^w L_0^{yw*} a_i^{sy}, a_i^{sy}\rangle \\ &+ \sum_{i=0}^{\infty}\langle [B^w R^w B^{w*} + B^u R^u B^{u*}]T^{*i}\xi_{N+1}, T^{*i}\xi_{N+1}\rangle \\ &+ \langle R^u(c_0^{su} - a_0^{su}), (c_0^{su} - a_0^{su})\rangle + \sum_{i=1}^{N}\langle R^u(B^{u*}\xi_i + L_0^{yu*}\nu_i + a_i^{su}), (B^{u*}\xi_i + L_0^{yu*}\nu_i + a_i^{su})\rangle \\ &+ \langle R^w c_0^{sw}, c_0^{sw}\rangle + \|v\|^2; \end{aligned} \tag{3.40}$$

i.e., when (3.11) holds and u is white,

$$\begin{aligned} J_{EN}(a^{sy}, a^{su}) &= J_{CN}(a^{sy}) + \langle R^w c_0^{sw}, c_0^{sw}\rangle + \|v\|^2 + \langle R^u(c_0^{su} - a_0^{su}), (c_0^{su} - a_0^{su})\rangle \\ &+ \sum_{i=1}^{N}\langle R^u[B^{u*}\xi_i + L_0^{yu*}a_i^{sy} + a_i^{su}], [B^{u*}\xi_i + L_0^{yu*}a_i^{sy} + a_i^{su}]\rangle. \quad \square \end{aligned} \tag{3.41}$$

Proof of Theorem 3.10 From (3.37) and (3.41). $\square$

Proof of Theorem 3.11 With $a^{sy} = 0$, (3.30) reduces to

$$\begin{aligned} J_E(0, a^{su}) &= \|A^{su}u - L^{su}u\|^2 + \|L^{sw}w\|^2 + \|v\|^2 \\ &= \|\sum_{i=0}^{\infty}[q^{-i}u, (c_i^{su} - a_i^{su})]\|^2 + \|L^{sw}w\|^2 + \|v\|^2. \end{aligned} \tag{3.42}$$

When u is white,

$$J_{EN}(0, a^{su}) = \sum_{i=0}^{N}\langle R^u(c_i^{su} - a_i^{su}), (c_i^{su} - a_i^{su})\rangle + \sum_{i=N+1}^{\infty}\langle R^u c_i^{su}, c_i^{su}\rangle + \|L^{sw}w\|^2 + \|v\|^2. \quad \square \tag{3.43}$$

4 Minimax Estimation and Prediction

From here on, we assume Hypothesis 3.1, the hypotheses of Problem 3.6, and (3.24) and (3.25) for $k \geq 1$. We also assume the following hypothesis.

Hypothesis 4.1 *There exist a Hilbert space Z and sequences $L_k^{zu} \in \mathcal{B}(U, Z)$ and $L_k^{zw} \in \mathcal{B}(W, Z)$. The sequence $L_0^{zw} w$ is independent of $L_k^{zw} w$ and $L_k^{sw} w$ $\forall k \geq 1$, and $L_0^{zw} w$ is independent of $L_k^{yw} w$ $\forall k \geq 0$. The sequence u is white.*

We define operators L^{zu} and L^{zw} as in (3.3). For each absolutely summable sequence $a_k^{sz} \in Z$, we define a corresponding sequence of linear functionals A_k^{sz} and an operator A^{sz} as in (3.1)–(3.3). This section concerns the linear system in (3.8) and (3.9) with the additional, Z-valued measurment sequence z satisfying

$$z = L^{zu} u + L^{zw} w. \tag{4.1}$$

In some applications, $z = s$.

We use the following notation for finite and infinite sequences

$$a_{j:k}^{sy} = [a_j^{sy}\ a_{j+1}^{sy}\ \dots a_k^{sy}], \qquad a_{j:k+1}^{sy} = [a_{j:k}^{sy}\ a_{k+1}^{sy}], \qquad a^{sy} = [a_0^{sy}\ a_{1:k}^{sy}\ a_{k+1}^{sy}\ a_{k+2}^{sy}\ \dots]. \tag{4.2}$$

For each positive integer N, we define the prediction-error sequence

$$\varepsilon\ =\ \sum_{k=0}^{N} A_k^{su} q^{-k} u + \sum_{k=1}^{N} A_k^{sy} q^{-k} y + \sum_{k=1}^{N} A_k^{sz} q^{-k} z\ -\ s, \tag{4.3}$$

and the fit-to-data criterion

$$J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz}) = \|\varepsilon\|^2 - \gamma^2 \sum_{k=1}^{N} |a_k^{sz}|^2, \tag{4.4}$$

where γ is a nonnegative real number.

Definition 4.2 *A real-valued function $f(\nu^y, \nu^u, \nu^z)$ is* coercive *in ν^y if*

$$f(\nu^y, \nu^u, \nu^z) = f_0 + f_1(\nu^y, \nu^u, \nu^z) + f_2(\nu^y, \nu^u, \nu^z) \tag{4.5}$$

where f_0 is a fixed real number, $f_1(\nu^y, \nu^u, \nu^z)$ is a linear function of (ν^y, ν^u, ν^z), and there exists a positive real number ρ such that

$$f_2(\nu^y, 0, 0) \geq \rho |\nu^y|^2. \tag{4.6}$$

If the corresponding condition holds for ν^u or ν^z, then f is coercive in ν^u or ν^z, respectively.

Problem 4.3 (Minimax: Finite Length ARX) *Suppose that $J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz})$ is coercive in $a_{1:N}^{sy}$ and $a_{0:N}^{su}$ and that $-J_{EN}$ is coercive in $a_{1:N}^{sz}$. Find sequences $a_{1:N}^{sy}$ in Y, $a_{0:N}^{su}$ in U, and $a_{1:N}^{sz}$ in Z that satisfy the saddle-point condition*

$$J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su}, \nu_{1:N}^{z}) \leq J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz}) \leq J_{EN}(\nu_{1:N}^{y}, \nu_{0:N}^{u}, a_{1:N}^{sz}) \tag{4.7}$$

for all sequences $\nu_{1:N}^z$ in Z, $\nu_{1:N}^y$ in Y, and $\nu_{0:N}^u$ in U.

Next we define

$$J_{EN,N}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz}) = J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz}). \tag{4.8}$$

For $1 \leq k < N$, when $J_{EN,k+1}([a_{1:k}^{sy}\ a_{k+1}^{sy}], [a_{0:k}^{su}\ a_{k+1}^{su}], [a_{1:k}^{sz}\ a_{k+1}^{sz}])$ is coercive in a_{k+1}^{sy} and a_{k+1}^{su} and $-J_{EN,k+1}$ is coercive in a_{k+1}^{sz} (for $a_{1:k}^{sy}$, $a_{0:k}^{su}$, and $a_{1:k}^{sz}$ fixed), we define

$$J_{EN,k}(a_{1:k}^{sy}, a_{0:k}^{su}, a_{1:k}^{sz}) = \min_{\substack{\eta^y \in Y \\ \eta^u \in U}} \max_{\eta^z \in Z} J_{EN,k+1}([a_{1:k}^{sy}\ \eta^y], [a_{0:k}^{su}\ \eta^u], [a_{1:k}^{sz}\ \eta^z]). \tag{4.9}$$

(According to (4.2), $a^{sy}_{1:k+1} = [a^{sy}_{1:k}\ a^{sy}_{k+1}]$, etc.) We also define

$$J_{EN,0}(0, a_0^{su}, 0) = \min_{\substack{\eta^y \in Y \\ \eta^u \in U}} \max_{\eta^z \in Z} J_{EN,1}(\eta^y, [a_0^{su}\eta^u], \eta^z). \tag{4.10}$$

Theorem 4.4 *If $J_{EN,k+1}([a^{sy}_{1:k}\ a^{sy}_{k+1}], [a^{su}_{0:k}\ a^{su}_{k+1}], [a^{sz}_{1:k}\ a^{sz}_{k+1}])$ is coercive in a^{sy}_{k+1} and a^{su}_{k+1} and $-J_{EN,k+1}$ is coercive in a^{sz}_{k+1} $\forall\ k = 0, 1, \ldots, N-1$, then for each such k there exist bounded linear operators $K^{yy}_{N,k}, K^{yu}_{N,k}, K^{yz}_{N,k}, K^{uy}_{N,k}, \ldots, K^{zz}_{N,k}$ and vectors $\phi^y_k, \phi^u_k, \phi^z_k$ such that*

$$J_{EN,k}(a^{sy}_{1:k}, a^{su}_{1:k}, a^{sz}_{1:k}) = J_{EN,k+1}([a^{sy}_{1:k}\ a^{sy}_{k+1}], [a^{su}_{0:k}\ a^{su}_{k+1}], [a^{sz}_{1:k}\ a^{sz}_{k+1}]) \tag{4.11}$$

with

$$a^{sy}_{k+1} = K^{yy}_{N,k}a^{sy}_{1:k} + K^{yu}_{N,k}a^{su}_{0:k} + K^{yz}_{N,k}a^{sz}_{1:k} + \phi^y_{k+1}, \tag{4.12}$$

$$a^{su}_{k+1} = K^{uy}_{N,k}a^{sy}_{1:k} + K^{uu}_{N,k}a^{su}_{0:k} + K^{uz}_{N,k}a^{sz}_{1:k} + \phi^u_{k+1}, \tag{4.13}$$

$$a^{sz}_{k+1} = K^{zy}_{N,k}a^{sy}_{1:k} + K^{zu}_{N,k}a^{su}_{0:k} + K^{zz}_{N,k}a^{sz}_{1:k} + \phi^z_{k+1}, \tag{4.14}$$

$$a^{sy}_{1:0} = 0, \qquad a^{sz}_{1:0} = 0. \tag{4.15}$$

Furthermore, for each k, the a^{sy}_{k+1}, a^{su}_{k+1}, and a^{sz}_{k+1} for which (4.11) *holds are unique.*

This theorem follows from the assumed coercivity and the fact that J_{EN} is a linear-quadratic functional.

Problem 4.5 (Order-recursive Minimax: Finite Length ARX) *Suppose that $J_{EN,k+1}([a^{sy}_{1:k}\ a^{sy}_{k+1}], [a^{su}_{0:k}\ a^{su}_{k+1}], [a^{sz}_{1:k}\ a^{sz}_{k+1}])$ is coercive in a^{sy}_{k+1} and a^{su}_{k+1} and $-J_{EN,k+1}$ is coercive in a^{sz}_{k+1} $\forall\ k = 0, 1, \ldots, N-1$. Find $a^{sy}_{1:N}$ and $a^{su}_{0:N}$ such that a_0^{su} minimizes $J_{EN,0}(0, a_0^{su}, 0)$ and*

$$a^{sy}_{k+1} = K^{yy}_{N,k}a^{sy}_{1:k} + K^{yu}_{N,k}a^{su}_{0:k} + \phi^y_{k+1}, \qquad 0 \le k \le N-1, \tag{4.16}$$

$$a^{su}_{k+1} = K^{uy}_{N,k}a^{sy}_{1:k} + K^{uu}_{N,k}a^{su}_{0:k} + \phi^u_{k+1}, \qquad 0 \le k \le N-1, \tag{4.17}$$

$$a^{sy}_{1:0} = 0. \tag{4.18}$$

Now we consider the optimal control problems of finding open-loop and closed-loop saddle-point strategies for the performance index

$$\begin{aligned} J_{CN}(\nu^y_{1:N}, \nu^z_{1:N}) = &\sum_{i=1}^{N} \langle R^w B^{w*}\xi_i, B^{w*}\xi_i\rangle + \langle G\xi_{N+1}, \xi_{N+1}\rangle \\ &+ \sum_{i=1}^{N} [\langle L_0^{yw} R^w L_0^{yw*}\nu^y_i, \nu^y_i\rangle + \langle L_0^{zw} R^w L_0^{zw*}\nu^z_i, \nu^z_i\rangle] - \gamma^2 \sum_{i=1}^{N} |\nu^z_i|^2 \end{aligned} \tag{4.19}$$

where G is the operator in (3.22) and the sequence ξ in X satisfies

$$\xi_1 = -c^{sx}, \qquad \xi_{k+1} = T^*\xi_k + C^{y*}\nu^y_k + C^{z*}\nu^z_k \quad \forall k \ge 1. \tag{4.20}$$

We use the standard notions of saddle-point strategies for linear-quadratic games.

Problem 4.6 (Open-loop LQ Game: Finite Interval) *Suppose that $J_{CN}(\nu^y_{1:N}, \nu^z_{1:N})$ is coercive in $\nu^y_{1:N}$ and $-J_{CN}$ is coercive in $\nu^z_{1:N}$. Find sequences $a^{sy}_{1:N}$ in Y and $a^{sz}_{1:N}$ in Z that satisfy the saddle point condition*

$$J_{CN}(a^{sy}_{1:N}, \nu^z_{1:N}) \le J_{CN}(a^{sy}_{1:N}, a^{sz}_{1:N}) \le J_{CN}(\nu^y_{1:N}, a^{sz}_{1:N}) \tag{4.21}$$

for all sequences $\nu^z_{1:N}$ in Z and $\nu^y_{1:N}$ in Y.

We will define closed-loop saddle-point strategies in a way that is most relevant to Problem 4.5, but our definition here is equivalent to the standard definition (see [6], for example) of closed-loop saddle-point strategies for (4.19) and (4.20). We define

$$J_{CN,N}(\nu^y_{1:N},\nu^z_{1:N}) = J_{CN}(\nu^y_{1:N},\nu^z_{1:N}). \tag{4.22}$$

For $1 \le k < N$, when $J_{CN,k+1}([\nu^y_{1:k}\ \nu^y_{k+1}],[\nu^z_{1:k}\ \nu^z_{k+1}])$ is coercive in ν^y_{k+1} and $-J_{CN,k+1}$ is coercive in ν^z_{k+1} (for $\nu^y_{1:k}$ and $\nu^z_{1:k}$ fixed), we define

$$J_{CN,k}(\nu^y_{1:k},\nu^z_{1:k}) = \min_{\eta^y\in Y}\max_{\eta^z\in Z} J_{CN,k+1}([\nu^y_{1:k}\ \eta^y],[\nu^z_{1:k}\ \eta^z]). \tag{4.23}$$

Theorem 4.7 *If $J_{CN,k+1}([\nu^y_{1:k}\ \nu^y_{k+1}],[\nu^z_{1:k}\ \nu^z_{k+1}])$ is coercive in ν^y_{k+1} and $-J_{CN,k+1}$ is coercive in ν^z_{k+1} $\forall$ $k = 0,1,\dots,N-1$, then for each such k there exist bounded linear operators F^y_{k+1} and F^z_{k+1} such that*

$$J_{CN,k}(\nu^y_{1:k},\nu^z_{1:k}) = J_{CN,k+1}([\nu^y_{1:k}\ \nu^y_{k+1}],[\nu^z_{1:k}\ \nu^z_{k+1}]) \tag{4.24}$$

with

$$\nu^y_{k+1} = -F^y_{k+1}\xi_{k+1}, \tag{4.25}$$

$$\nu^z_{k+1} = -F^z_{k+1}\xi_{k+1}. \tag{4.26}$$

Also, the unique sequences $a^{*y}_{1:N}$ and $a^{*z}_{1:N}$ that solve Problem 4.6 *are $a^{*y}_{1:N} = \nu^y_{1:N}$ and $a^{*z}_{1:N} = \nu^z_{1:N}$ with $\nu^y_{1:N}$ and $\nu^z_{1:N}$ generated by* (4.25) *and* (4.26).

This theorem follows from the assumed coercivity and the fact that J_{CN} is a linear-quadratic functional. The feedback control laws in (4.25) and (4.26) are the pair of *closed-loop saddle-point strategies* for (4.19) and (4.20).

Our formulation of the order-recursive minimax parameter-estimation problem and the closed-loop saddle-point strategies for the control problem are motivated by the following standard results on LQ dynamic games [6]. There exists a sequence of Riccati operators P_k, generated by a Riccati difference equation backward from the final condition $P_{N+1} = G$, such that

$$\begin{aligned} J_{CN,k}(\nu^y_{1:k},\nu^z_{1:k}) &= \sum_{i=1}^{k}\langle R^w B^{w*}\xi_i, B^{w*}\xi_i\rangle + \langle P_{k+1}\xi_{k+1},\xi_{k+1}\rangle \\ &+ \sum_{i=1}^{k}[\langle L_0^{yw} R^w L_0^{yw*}\nu^y_i,\nu^y_i\rangle + \langle L_0^{zw} R^w L_0^{zw*}\nu^z_i,\nu^z_i\rangle] - \gamma^2\sum_{i=1}^{k}|\nu^z_i|^2 \end{aligned} \tag{4.27}$$

for any sequences $\nu^y_{1:k}$ and $\nu^z_{1:k}$. The gain operators F^y_{k+1} and F^z_{k+1} in (4.25) and (4.26) can be constructed from P_{k+1} and the operators in (4.19) and (4.20). See Section 5.2. The standard results on optimal control do not appear to provide an algorithm for solving the parameter-estimation problem, but they make the solutions to Problems 4.6 and 4.8 trivial.

Problem 4.8 (Closed-loop LQ Game: Finite Interval) *Suppose that the coercivity hypotheses in* Theorem 4.7 *hold. Find the sequence $\nu^y_{1:N}$ generated by*

$$\nu^y_k = -F^y_k\xi_k, \tag{4.28}$$

and (4.20) *with $\nu^z_{1:N} = 0$.*

Theorem 4.9 *Suppose that the coercivity hypotheses in* Problem 4.3 *hold. Then the coercivity hypotheses in* Problem 4.6 *hold, and the finite sequences $a^{*y}_{1:N}$, $a^{*u}_{0:N}$ and $a^{*z}_{1:N}$ solve* Problem 4.3 *if and only if $a^{*y}_{1:N}$ and $a^{*z}_{1:N}$ solve* Problem 4.6 *and*

$$a^{*u}_i = -B^{u*}\xi_i - L_0^{yu*}a^{*y}_i - L_0^{zu*}a^{*z}_i, \qquad i \ge 1, \qquad a^{*u}_0 = c^{*u}_0. \tag{4.29}$$

Theorem 4.10 *Suppose that the coercivity hypotheses in* Problem 4.5 *hold. Then the coercivity hypotheses in* Theorem 4.7 *hold, and the finite sequences* $a^{*y}_{1:N}$ *and* $a^{*u}_{0:N}$ *solve* Problem 4.5 *if and only if* $\nu^{y}_{1:N} = a^{*y}_{1:N}$ *solves* Problem 4.8 *and*

$$a^{*u}_i = -B^{u*}\xi_i - L_0^{yu*}a^{*y}_i, \qquad i \geq 1, \qquad a^{*u}_0 = c^{*u}_0. \tag{4.30}$$

Proofs of Theorems 4.9–4.16 In (3.30)–(3.41), we include terms corresponding to the measurement z in (4.1) to obtain

$$\begin{aligned}J_{EN}(a^{*y}_{1:N}, a^{*u}_{0:N}, a^{*z}_{1:N}) &= J_{CN}(a^{*y}_{1:N}, a^{*z}_{1:N}) + \langle R^u(c^{*u}_0 - a^{*u}_0), (c^{*u}_0 - a^{*u}_0)\rangle \\ &+ \sum_{i=1}^{N}\langle R^u[B^{u*}\xi_i + L_0^{yu*}a^{*y}_i + L_0^{zu*}a^{*z}_i + a^{*u}_i], [B^{u*}\xi_i + L_0^{yu*}a^{*y}_i + L_0^{zu*}a^{*z}_i + a^{*u}_i]\rangle \\ &+ \langle R^w c^{*w}_0, c^{*w}_0\rangle + \|v\|^2\end{aligned} \tag{4.31}$$

where $J_{CN}(a^{*y}_{1:N}, a^{*z}_{1:N})$ is given by (4.19), ξ satisfies (4.20) with $\nu^{y}_{1:N} = a^{*y}_{1:N}$ and $\nu^{z}_{1:N} = a^{*z}_{1:N}$, and $\langle R^w c^{*w}_0, c^{*w}_0\rangle$ and $\|v\|^2$ are constants independent of $a^{*y}_{1:N}$, $a^{*u}_{0:N}$, and $a^{*z}_{1:N}$. The following lemma then implies Theorems 4.9 and 4.10. The subsequent Theorems 4.15 and 4.16 follow from (4.31).

Lemma 4.11 *Let H_i ($i = 1, 2, \ldots, 5$) be Hilbert spaces with $L_{ij} \in \mathcal{B}(H_i, H_j)$, $c_4 \in H_4$, $c_5 \in H_5$. Define $J(\cdot,\cdot) : H_1 \times H_3 \to \mathbf{R}$ and $\hat{J}(\cdot,\cdot) : H_1 \times H_2 \times H_3 \to \mathbf{R}$ by*

$$J(h_1, h_3) = |L_{41}h_1 + L_{43}h_3 + c_4|^2 - \gamma^2|h_3|^2, \tag{4.32}$$

$$\hat{J}(h_1, h_2, h_3) = J(h_1, h_3) + |L_{51}h_1 + h_2 + L_{53}h_3 + c_5|^2, \tag{4.33}$$

where γ is a positive real number, and consider the saddle-point conditions

$$J(h_1, \tilde{h}_3) \leq J(h_1, h_3) \leq J(\tilde{h}_1, h_3) \qquad \forall\, \tilde{h}_1 \in H_1,\ \tilde{h}_3 \in H_3, \tag{4.34}$$

$$\hat{J}(\hat{h}_1, \hat{h}_2, \tilde{h}_3) \leq \hat{J}(\hat{h}_1, \hat{h}_2, \hat{h}_3) \leq \hat{J}(\tilde{h}_1, \tilde{h}_2, \hat{h}_3) \qquad \forall\, \tilde{h}_1 \in H_1,\ \tilde{h}_2 \in H_2,\ \tilde{h}_3 \in H_3. \tag{4.35}$$

Suppose that (4.35) *holds for some $(\hat{h}_1, \hat{h}_2, \hat{h}_3) \in H_1 \times H_2 \times H_3$. Then*

$$\hat{h}_2 = -(L_{51}\hat{h}_1 + L_{53}\hat{h}_3 + c_5), \tag{4.36}$$

and the following one-to-one correspondence exists between saddle points of J and saddle points of $\hat{J}$: (4.34) *holds for some $(h_1, h_3) \in H_1 \times H_3$ if and only if* (4.35) *holds for*

$$\hat{h}_1 = h_1, \qquad \hat{h}_2 = -(L_{51}h_1 + L_{53}h_3 + c_5), \qquad \hat{h}_3 = h_3. \tag{4.37}$$

This lemma follows from the first-order necessary conditions (normal equations) for saddle points of linear-quadratic functionals and the fact that J is concave in h_3 if $\hat{J}$ is.

Theorems 4.9 and 4.10 say that Problems 4.3 and 4.5 are equivalent, respectively, to Problems 4.6 and 4.8 when γ is large enough for the appropriate objective functionals in Problems 4.3 and 4.5 to be coercively concave (i.e., $-J$ coercive) in the maximizing parameters $a^{*z}_{1:N}$. Because of the term on the second line of (4.31), γ generally must be larger for the coercivity hypotheses in Problems 4.3 and 4.5 than for the coercivity hypotheses in Problems 4.6 and 4.8. The following alternative minimax parameter estimation problems have coercivity hypotheses that are equivalent to those in the corresponding control problems.

We define

$$\tilde{J}_{EN,N}(a^{*y}_{1:N}, a^{*z}_{1:N}) = \tilde{J}_{EN}(a^{*y}_{1:N}, a^{*z}_{1:N}) = \min_{a_{0:N}} J_{EN}(a^{*y}_{1:N}, a_{0:N}, a^{*z}_{1:N}). \tag{4.38}$$

Problem 4.12 (Minimax: Finite Length ARX) *Suppose that* $\tilde{J}_{EN}(a^{*y}_{1:N}, a^{*z}_{1:N})$ *is coercive in* $a^{*y}_{1:N}$ *and that* $-\tilde{J}_{EN}$ *is coercive in* $a^{*z}_{1:N}$. *Find sequences* $a^{*y}_{1:N}$ *in* Y *and* $a^{*z}_{1:N}$ *in* Z *that satisfy the saddle-point condition*

$$\tilde{J}_{EN}(a^{*y}_{1:N}, \nu^{z}_{1:N}) \leq \tilde{J}_{EN}(a^{*y}_{1:N}, a^{*z}_{1:N}) \leq \tilde{J}_{EN}(\nu^{y}_{1:N}, a^{*z}_{1:N}) \tag{4.39}$$

for all sequences $\nu^{z}_{1:N}$ *in* Z *and* $\nu^{y}_{1:N}$ *in* Y, *and for this* $a^{*y}_{1:N}$ *and* $a^{*z}_{1:N}$, *find* $a^{*u}_{0:N}$ *such that*

$$\tilde{J}_{EN}(a^{*y}_{1:N}, a^{*z}_{1:N}) = J_{EN}(a^{*y}_{1:N}, a^{*u}_{0:N}, a^{*z}_{1:N}). \tag{4.40}$$

For $1 \leq k < N$, when $\tilde{J}_{EN,k+1}([a^{*y}_{1:k}\ a^{*y}_{k+1}], [a^{*z}_{1:k}\ a^{*z}_{k+1}])$ is coercive in a^{*y}_{k+1} and $-\tilde{J}_{EN,k+1}$ is coercive in a^{*z}_{k+1}, we define

$$\tilde{J}_{EN,k}(a^{*y}_{1:k}, a^{*z}_{1:k}) = \min_{\eta^y \in Y} \max_{\eta^z \in Z} \tilde{J}_{EN,k+1}([a^{*y}_{1:k}\ \eta^y], [a^{*z}_{1:k}\ \eta^z]). \tag{4.41}$$

Theorem 4.13 *If* $\tilde{J}_{EN,k+1}([a^{*y}_{1:k}\ a^{*y}_{k+1}], [a^{*z}_{1:k}\ a^{*z}_{k+1}])$ *is coercive in* a^{*y}_{k+1} *and* $-\tilde{J}_{EN,k+1}$ *is coercive in* a^{*z}_{k+1} $\forall\ k = 0, 1, \ldots, N-1$, *then for each such* k *there exist bounded linear operators* $\tilde{K}^{yy}_{N,k}, \tilde{K}^{yz}_{N,k}, \tilde{K}^{zy}_{N,k}, \tilde{K}^{zz}_{N,k}$ *and vectors* $\tilde{\phi}^y_k, \tilde{\phi}^z_k$ *such that*

$$\tilde{J}_{EN,k}(a^{*y}_{1:k}, a^{*z}_{1:k}) = \tilde{J}_{EN,k+1}([a^{*y}_{1:k}\ a^{*y}_{k+1}], [a^{*z}_{1:k}\ a^{*z}_{k+1}]) \tag{4.42}$$

with

$$a^{*y}_{k+1} = \tilde{K}^{yy}_{N,k} a^{*y}_{1:k} + \tilde{K}^{yz}_{N,k} a^{*z}_{1:k} + \tilde{\phi}^y_{k+1}, \tag{4.43}$$

$$a^{*z}_{k+1} = \tilde{K}^{zy}_{N,k} a^{*y}_{1:k} + \tilde{K}^{zz}_{N,k} a^{*z}_{1:k} + \tilde{\phi}^z_{k+1}, \tag{4.44}$$

$$a^{*y}_{1:0} = 0, \qquad a^{*z}_{1:0} = 0. \tag{4.45}$$

Furthermore, for each k, *the* a^{*y}_{k+1} *and* a^{*z}_{k+1} *for which* (4.42) *holds are unique.*

This theorem, like Theorem 4.4, follows from the assumed coercivity and the fact that $\tilde{J}_{EN}$ is a linear-quadratic functional.

Problem 4.14 (Order-recursive Minimax: Finite Length ARX) *Suppose that* $\tilde{J}_{EN,k+1}([a^{*y}_{1:k}\ a^{*y}_{k+1}], [a^{*z}_{1:k}\ a^{*z}_{k+1}])$ *is coercive in* a^{*y}_{k+1} *and* $-\tilde{J}_{EN,k+1}$ *is coercive in* a^{*z}_{k+1} $\forall\ k = 0, 1, \ldots, N-1$. *Find* $a^{*y}_{1:N}$ *such that*

$$a^{*y}_{k+1} = \tilde{K}^{yy}_{N,k} a^{*y}_{1:k} + \tilde{\phi}^y_{k+1}, \qquad 0 \leq k \leq N-1, \tag{4.46}$$

$$a^{*y}_{1:0} = 0, \tag{4.47}$$

and for this $a^{*y}_{1:N}$, *find* $a^{*u}_{0:N}$ *such that*

$$\tilde{J}_{EN}(a^{*y}_{1:N}, 0) = J_{EN}(a^{*y}_{1:N}, a^{*u}_{0:N}, 0). \tag{4.48}$$

Theorem 4.15 *The coercivity hypotheses in* Problem 4.12 *hold if and only if the coercivity hypotheses in* Probl m 4.6 *hold. When these hypotheses hold, the finite sequences* $a^{*y}_{1:N}$, $a^{*u}_{0:N}$ *and* $a^{*z}_{1:N}$ *solve* Problem 4.12 *if and only if* $a^{*y}_{1:N}$ *and* $a^{*z}_{1:N}$ *solve* Problem 4.6 *and* $a^{*u}_{0:N}$ *is given by* (4.29).

Theorem 4.16 *The coercivity hypotheses in* Problem 4.14 *hold if and only if the coercivity hypotheses in* Theorem 4.7 *hold. When these hypotheses hold, the finite sequences* $a^{*y}_{1:N}$ *and* $a^{*u}_{0:N}$ *solve* Problem 4.14 *if and only if* $\nu^{y}_{1:N} = a^{*y}_{1:N}$ *solves* Problem 4.8 *and* $a^{*u}_{0:N}$ *is given by* (4.30).

5 Relationships to State-Space Filtering

5.1 LQR/Kalman Filter Representations of LS Parameters

We begin with the finite-order LS problem and the finite-time control and filtering problems. According to standard results on LQR problems, the solution to Problem 3.7 has the form

$$\nu_k = -F_k \xi_k \tag{5.1}$$

where

$$F_k = [L_0^{yw} R^w L_0^{yw*} + C^y P_{k+1} C^{y*}]^{-1} C^y P_{k+1} T^* \tag{5.2}$$

and the Riccati operators P_k are the nonnegative, self-adjoint elements of $\mathcal{B}(X, X)$ satisfying

$$P_k = B^w R^w B^{w*} + T(P_{k+1} - P_{k+1} C^{y*} [L_0^{yw} R^w L_0^{yw*} + C^y P_{k+1} C^{y*}]^{-1} C^y P_{k+1}) T^*, \tag{5.3}$$

$$P_{N+1} = G. \tag{5.4}$$

The corresponding optimal sequence ξ satisfies

$$\xi_k = \tilde{T}_{k-1}^* \tilde{T}_{k-2}^* \cdots \tilde{T}_0^* (-c^{sx}) \qquad 1 \le k \le N+1, \tag{5.5}$$

where

$$\tilde{T}_k = T - F_k^* C^y, \qquad 1 \le k \le N, \qquad \qquad \tilde{T}_0 = I. \tag{5.6}$$

With (5.1) and (5.5), the solution to Problem 3.3 given in Theorem 3.9 becomes

$$a_k^{sy} = F_k \tilde{T}_{k-1}^* \tilde{T}_{k-2}^* \cdots \tilde{T}_0^* c^{sx}, \qquad 1 \le k \le N, \tag{5.7}$$

$$a_k^{su} = B^{u*} \tilde{T}_{k-1}^* \tilde{T}_{k-2}^* \cdots \tilde{T}_0^* c^{sx} - L_0^{yu*} a_k^{sy}, \qquad 1 \le k \le N, \qquad a_0^{su} = c_0^{su}. \tag{5.8}$$

According to (3.1) and (3.17), the bounded linear functionals corresponding to the a_k^{sy} and a_k^{su} in (5.7) and (5.8) are

$$A_k^{sy} = C^s \tilde{T}_0 \cdots \tilde{T}_{k-2} \tilde{T}_{k-1} F_k^*, \qquad 1 \le k \le N, \tag{5.9}$$

$$A_k^{su} = C^s \tilde{T}_0 \cdots \tilde{T}_{k-2} \tilde{T}_{k-1} B^u - A_k^{sy} L_0^{yu} \qquad 1 \le k \le N, \qquad A_0^{su} = L_0^{su}. \tag{5.10}$$

Under the hypotheses of Theorems 3.8 and 3.9, a state-space realization of the input/output system in (3.8) and (3.9) is

$$x_{k+1} = T x_k + B^u u_k + B^w w_k, \tag{5.11}$$

$$y_k = C^y x_k + L_0^{yu} u_k + L_0^{yw} w_k, \tag{5.12}$$

$$s_k = C^s x_k + L_0^{su} u_k + L_0^{sw} w_k + v. \tag{5.13}$$

(Recall that u is a known input sequence while w and v are unknown noise sequences.) If y is the measured output, a Kalman filter for one-step-ahead prediction of x_{k+1} has the form

$$\hat{x}_{k+1} = \hat{T}_k \hat{x}_k + \hat{F}_k y_k + [B^u - \hat{F}_k L_0^{yu}] u_k, \tag{5.14}$$

and the corresponding one-step-ahead prediction of s_{k+1} is

$$\hat{s}_{k+1} = C^s \hat{x}_{k+1} + L_0^{su} u_{k+1} = \langle \hat{x}_{k+1}, c^{sx} \rangle + L_0^{su} u_{k+1}. \tag{5.15}$$

(The $\hat{x}_k$ here often is denoted by $\hat{x}_{k/k-1}$. See [20], for example.) If the expected value of x_1 is $\hat{x}_1 = 0$, then

$$\begin{aligned} \hat{x}_{N+1} &= \hat{F}_N y_N + [B^u - \hat{F}_N L_0^{yu}] u_N \\ &+ \sum_{i=1}^{N-1} \hat{T}_N \hat{T}_{N-1} \cdots \hat{T}_{N+1-i} (\hat{F}_{N-i} y_{N-i} + [B^u - \hat{F}_{N-i} L_0^{yu}] u_{N-i}). \end{aligned} \tag{5.16}$$

If the probabilistic error-covariance operator for $\hat{x}_1$ is G, then

$$\hat{F}_k = F^*_{N+1-k}, \qquad \hat{T}_k = \tilde{T}_{N+1-k}, \qquad 1 \le k \le N, \tag{5.17}$$

where F_k and $\tilde{T}_k$ are the operators in (5.1)–(5.6). Substituting (5.16) into (5.15) and using (5.17), (5.7), and (5.8) yield

$$\begin{aligned}\hat{s}_{N+1} &= \sum_{i=1}^{N} \langle y_{N+1-i}, a_i^{sy} \rangle + \sum_{i=0}^{N} \langle u_{N+1-i}, a_i^{su} \rangle \\ &= \sum_{i=1}^{N} A_i^{sy} y_{N+1-i} + \sum_{i=0}^{N} A_i^{su} u_{N+1-i}.\end{aligned} \tag{5.18}$$

Thus the ARX coefficients that minimize the least-squares one-step-ahead prediction error in Problem 3.3 are the coefficients in the probabilistic minimum-variance prediction of $\hat{s}_{N+1}$ based on the data $y_1, y_2, \dots y_N$ and $u_1, u_2, \dots u_N$ and the assumption that the initial state vector x_1 has zero mean and covariance G, given by (3.22). This G is indeed correct for the steady-state statistics of the state vector x_k in (5.11) when T has spectral radius less than 1 and u_k and w_k are zero-mean stationary white noise sequences (in the probalistic sense) with covariance operators R^u and R^w, respectively.

When we say that the parameters a_i^{sy} and a_i^{su}, or equivalently A_i^{sy} and A_i^{su}, are Markov parameters for a state-space filter, we mean that they are the coefficients in a prediction formula like (5.18).

Now we turn to the infinite-order LS problem and the infinite-time control and filtering problems. The solution to Problem 3.7 is given by (5.1)–(5.6) with F_k, P_k, and $\tilde{T}_k$ independent of k. Hence

$$F_k = F, \qquad P_k = P, \qquad \tilde{T}_k = \tilde{T} = T - F^* C^y, \tag{5.19}$$

and the solution to Problem 3.2 given in Theorem 3.8 becomes

$$a_k^{sy} = F\tilde{T}^{*(k-1)} c^{sx}, \qquad k \ge 1, \tag{5.20}$$

$$a_k^{su} = B^{u*}\tilde{T}^{*(k-1)} c^{sx} - L_0^{yu*} a_k^{sy} \qquad k \ge 1, \qquad a_0^{su} = c_0^{su}. \tag{5.21}$$

For the corresponding bounded linear functionals, (5.9), (5.10), and (5.17) yield

$$A_k^{sy} = C^s \hat{T}^{(k-1)} \hat{F}, \qquad k \ge 1, \tag{5.22}$$

$$A_k^{su} = C^s \hat{T}^{(k-1)} B^u - A_k^{sy} L_0^{yu}, \qquad k \ge 1, \qquad A_0^{su} = L_0^{su}. \tag{5.23}$$

The linear functionals A_k^{su} and A_k^{sy} in (5.22) and (5.23) should be recognized easily as the Markov parameters for the steady-state Kalman filter for the system in (5.11)–(5.13) with measured output y and predicted output s.

5.2 LQ Game/H_∞ Filter Representation of Minimax Parameters

According to standard results on LQ games (see [6] for example), the solution to Problems 4.6 and 4.8 have the form

$$a_k^{*y} = \nu_k^y = -F_k^y \xi_k, \tag{5.24}$$

$$a_k^{*z} = \nu_k^z = -F_k^z \xi_k, \tag{5.25}$$

where

$$\begin{bmatrix} F_k^y \\ F_k^z \end{bmatrix} = [R + CP_{k+1}C^*]^{-1} CP_{k+1}T^*, \tag{5.26}$$

$$R = \begin{bmatrix} L_0^{yw} R^w L_0^{yw*} & 0 \\ 0 & (L_0^{zw} R^w L_0^{zw*} - \gamma^2 I) \end{bmatrix}, \qquad C = \begin{bmatrix} C^y \\ C^z \end{bmatrix}, \tag{5.27}$$

and the Riccati operators P_k are the nonnegative, self-adjoint elements of $\mathcal{B}(X, X)$ satisfying

$$P_k = B^w R^w B^{w*} + T(P_{k+1} - P_{k+1}C^*[R + CP_{k+1}C^*]^{-1}CP_{k+1})T^* \tag{5.28}$$

and the final condition (5.4). The corresponding optimal sequence ξ satisfies (5.5) with (5.6) replaced by

$$\tilde{T}_k = T - F_k^{y*}C^y - F_k^{z*}C^z, \qquad 1 \le k \le N, \qquad \tilde{T}_0 = I. \tag{5.29}$$

The coercivity hypotheses in Theorem 4.7 hold if and only if the operators $[L_0^{yw} R^w L_0^{yw*} + C^y P_{k+1} C^{y*}]$ and $[\gamma^2 I - L_0^{zw} R^w L_0^{zw*} - C^z P_{k+1} C^{z*}]$ are coercive for $1 \le k \le N$.

With (5.24), (5.25), (5.5), and (5.29), the solution to Problems 4.3 and 4.12 given in Theorems 4.9 and 4.15 becomes

$$a_k^{*y} = F_k^y \tilde{T}_{k-1}^* \tilde{T}_{k-2}^* \cdots \tilde{T}_0^* c^{*x}, \qquad 1 \le k \le N, \tag{5.30}$$

$$a_k^{*u} = B^{u*} \tilde{T}_{k-1}^* \tilde{T}_{k-2}^* \cdots \tilde{T}_0^* c^{*x} - L_0^{yu*} a_k^{*y} - L_0^{zu*} a_k^{*z}, \qquad 1 \le k \le N, \qquad a_0^{*u} = c_0^{*u}, \tag{5.31}$$

$$a_k^{*z} = F_k^z \tilde{T}_{k-1}^* \tilde{T}_{k-2}^* \cdots \tilde{T}_0^* c^{*x}, \qquad 1 \le k \le N. \tag{5.32}$$

The corresponding bounded linear functionals are

$$A_k^{*y} = C^* \tilde{T}_0 \cdots \tilde{T}_{k-2} \tilde{T}_{k-1} F_k^{y*}, \qquad 1 \le k \le N, \tag{5.33}$$

$$A_k^{*u} = C^* \tilde{T}_0 \cdots \tilde{T}_{k-2} \tilde{T}_{k-1} B^u - A_k^{*y} L_0^{yu} - A_k^{*z} L_0^{zu} \qquad 1 \le k \le N, \qquad A_0^{*u} = L_0^{*u}, \tag{5.34}$$

$$A_k^{*z} = C^* \tilde{T}_0 \cdots \tilde{T}_{k-2} \tilde{T}_{k-1} F_k^{z*}, \qquad 1 \le k \le N. \tag{5.35}$$

By the standard results on LQ games, the operators F_k^y in Theorem 4.7 and Problem 4.8 are those in (5.25)–(5.29). Hence, by Theorems 4.10 and 4.16, the sequences $a_{1:N}^{*y}$ and $a_{0:N}^{*u}$ in the solution to Problems 4.5 and 4.14 are given by (5.30) and (5.31) with $a_{1:N}^{*z} = 0$ and

$$\tilde{T}_k = T - F_k^{y*} C^y, \qquad 1 \le k \le N, \qquad \tilde{T}_0 = I. \tag{5.36}$$

The role of LQ game theory in H_∞ control and filtering stems from the following well known result, which is a corollary to Theorem 4.7.

Corollary 5.1 *Suppose that the coercivity hypotheses in* Theorem 4.7 *hold, that $\nu_{1:N}^y$ is generated by the minimizing closed-loop saddle-point strategy in* (4.25) *and* (5.24), *and that $\xi_1 = 0$. Then*

$$J_{CN}(\nu_{1:N}^y, \nu_{1:N}^z) \le 0 \qquad \forall\, \nu_{1:N}^z. \tag{5.37}$$

To interpret (5.37) in the most useful way for this paper, we assume that G can be factored as

$$G = \bar{B}\bar{B}^* \tag{5.38}$$

where $\bar{B}$ is a bounded linear operator from a Hilbert space $\bar{X}$ to X, and we define

$$\bar{\xi} = \bar{B}^*\xi_{N+1} \tag{5.39}$$

and

$$\eta_k = R^{w/2}(B^{w*}\xi_k + L_0^{yw*}\nu_k^y), \qquad 1 \le k \le N, \tag{5.40}$$

where $R^{w/2}$ is the nonnegative self-adjoint square root of R^w. Then,

$$\sum_{i=1}^{N}\langle R^w B^{w*}\xi_i, B^{w*}\xi_i\rangle + \langle G\xi_{N+1},\xi_{N+1}\rangle + \sum_{i=1}^{N}\langle L_0^{yw}R^w L_0^{yw*}\nu_i^y, \nu_i^y\rangle = \sum_{i=1}^{N}|\eta_i|^2 + |\bar{\xi}|^2. \tag{5.41}$$

If there exists a real number ρ such that

$$L_0^{zw}R^w L_0^{zw*} \ge \rho, \qquad \gamma > \rho, \tag{5.42}$$

then (5.37) yields

$$\sum_{i=1}^{N}|\eta_i|^2 + |\bar{\xi}|^2 \le (\gamma^2 - \rho^2)\sum_{i=1}^{N}|\nu_i^z|^2. \tag{5.43}$$

Under the hypotheses of Section 4, a state-space realization of the input/output system in (3.8), (3.9), and (4.1) is (5.11)-(5.13) with the additional output equation

$$z_k = C^z x_k + L_0^{zu}u_k + L_0^{zw}w_k. \tag{5.44}$$

If y is the measured output, a finite-time H_∞ filter for one-step-ahead prediction of x_{k+1} has the form (5.14) and the corresponding one-step-ahead prediction of s_{k+1} is given by (5.15), but the Kalman and H_∞ filters use different operators $\hat{T}_k$ and $\hat{F}_k$.

For the H_∞ filter that is related to the order-recursive minimax parameter estimation problems in Section 4, the operators $\hat{T}_k$ and $\hat{F}_k$ are given by (5.17) with F_k and $\tilde{T}_k$ generated by (5.26)-(5.28) and (5.36). That these operators $\hat{F}_k$ and $\hat{T}_k$ indeed yield an H_∞ filter will be established in Theorem 5.2. But whatever the meaning of the state-space filter in (5.14) and (5.15) for this choice of $\hat{F}_k$ and $\hat{T}_k$, the one-step-ahead prediction of s_{N+1} by this filter, when $\hat{x}_1 = 0$, can be written as in (5.18) with the coefficients a_i^{sy} and a_i^{su} given by (5.30) and (5.31) with $a_{1:N}^{sz} = 0$ and $\tilde{T}_k$ given by (5.36)—and these are the same parameters a_i^{sy} and a_i^{su} that solve Problems 4.5 and 4.14.

Theorem 5.2 *Suppose that the coercivity hypotheses in* Theorem 4.7 *hold and that* (5.42) *holds. Assume*

$$x_1 = \bar{B}\bar{x}, \qquad \hat{x}_1 = \bar{B}\hat{\bar{x}}, \tag{5.45}$$

$$w_k = R^{w/2}\bar{w}_k, \qquad 1 \le k \le N, \tag{5.46}$$

with $\bar{x}, \hat{\bar{x}} \in \bar{X}$ *and* $\bar{w}_k \in W$. *Let* $\hat{x}_{2:N}$ *be generated by* (5.14) *with* $\hat{T}_k$ *and* $\hat{F}_k$ *generated by* (5.17), (5.26)-(5.28), *and* (5.36), *and define the prediction of* z *by*

$$\hat{z}_k = C^z\hat{x}_k + L_0^{zu}u_k, \qquad 1 \le k \le N. \tag{5.47}$$

Then

$$\sum_{k=1}^{N}|\hat{z}_k - z_k|^2 \le (\gamma^2 - \rho^2)\Big(\sum_{k=1}^{N}|\bar{w}_k|^2 + |\hat{\bar{x}} - \bar{x}|^2\Big). \tag{5.48}$$

Theorem 5.2 follows from (5.43) and the following lemma.

Lemma 5.3 (Duality between Control and State Estimation) *Let X, W, Z, and $\bar{X}$ be Hilbert spaces with $S_k \in \mathcal{B}(X, X)$, $B_k \in \mathcal{B}(Z, X)$, $C_k \in \mathcal{B}(X, W)$, $D_k \in \mathcal{B}(Z, W)$ $(1 \le k \le N)$ and $\bar{C} \in \mathcal{B}(W, Z)$. Define*

$$\begin{aligned} \hat{B}_k &= C^*_{N+1-k}, \quad \hat{C}_k = B^*_{N+1-k}, \\ \hat{D}_k &= D^*_{N+1-k}, \quad \hat{S}_k = S^*_{N+1-k}, \\ & 1 \le k \le N. \end{aligned} \tag{5.49}$$

Consider the two systems

$$\begin{aligned} \xi_{k+1} &= S_k \xi_k + B_k \nu_k, \quad 1 \le k \le N, \\ \eta_k &= C_k \xi_k + D_k \nu_k, \quad 1 \le k \le N, \\ \bar{\xi} &= \bar{C} \xi_{N+1}, \qquad \xi_1 = 0, \end{aligned} \tag{5.50}$$

and

$$\begin{aligned} \xi_{k+1} &= \hat{S}_k \xi_k + \hat{B}_k \eta_k, \quad 1 \le k \le N, \\ \nu_k &= \hat{C}_k \xi_k + \hat{D}_k \eta_k, \quad 1 \le k \le N, \\ \xi_1 &= \bar{C}^* \bar{\xi}. \end{aligned} \tag{5.51}$$

(The ξ_k, $\bar{\xi}$, η_k, ν_k in (5.50) have no relation to the ξ_k, $\bar{\xi}$, η_k, ν_k in (5.51), except for belonging to the same respective spaces.) Let $L \in \mathcal{B}(Z \times Z \times \cdots \times Z, W \times W \times \cdots \times W \times \bar{X})$ be the operator such that, for the system in (5.50),

$$(\eta_{1:N}, \bar{\xi}) = L\, \nu_{1:N} \tag{5.52}$$

and let $\hat{L} \in \mathcal{B}(W \times W \times \cdots \times W \times \bar{X}, Z \times Z \times \cdots \times Z)$ be the operator such that, for the system in (5.51),

$$\nu_{1:N} = \hat{L}\,(\eta_{1:N}, \bar{\xi}). \tag{5.53}$$

Also, let I^η and I^ν be the isometric isomorphisms such that

$$I^\eta([\eta_1\ \eta_2\ \cdots\ \eta_N], \bar{\xi}) = ([\eta_N\ \cdots\ \eta_2\ \eta_1], \bar{\xi}), \qquad I^\nu[\nu_1\ \nu_2\ \cdots\ \nu_N] = [\nu_N\ \cdots\ \nu_2\ \nu_1]. \tag{5.54}$$

Then

$$\hat{L} = I^\nu L^* I^\eta. \tag{5.55}$$

Calculation of the adjoint operator L^* yields Lemma 5.3.

Proof of Theorem 5.2 Let (5.50) be the closed-loop control system in the closed-loop LQ game (Problem 4.8) and let (5.51) be the prediction-error system for the filter specified in Theorem 5.2. Then (5.48) follows from (5.43) and the fact that L and L^* have the same norm. (Related arguments appear to be used in [6].) □

6 Problems with Finite Amounts of Data: Convergence

\.. assume Hypothesis 3.1, the hypotheses of Problem 3.6, (3.24) and (3.25) for $k \geq 1$, and Hypothesis 4.1. In this section, we assume also that the input space U and the output spaces Y and Z are finite-dimensional. We do *not* assume that the noise space W or the state space X introduced in Problem 3.6 are finite-dimensional.

6.1 Least-Squares Estimation and Prediction

For each $t \geq 0$, we define

$$J_{EN}^{(t)}(a_{1:N}^{sy}, a_{0:N}^{su}) = \frac{1}{t}\sum_{k=0}^{t} |\varepsilon_k(t)|^2 \tag{6.1}$$

where

$$\varepsilon_k(t) \;=\; \sum_{i=0}^{N}\langle u_{k-i}, a_i^{su}\rangle + \sum_{i=1}^{N}\langle y_{k-i}, a_i^{sy}\rangle \;-\; s_k, \qquad 0 \leq k \leq t, \tag{6.2}$$

with $y_i = 0$ and $u_i = 0$ for $i < 0$. In this section, we write $J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su})$ instead of $J_{EN}(a^{sy}, a^{su})$ for the objective functional in Problem 3.3.

Problem 6.1 *For $t \geq 0$, find $a_{1:N}^{sy}(t)$ and $a_{0:N}^{su}(t)$ to minimize $J_{EN}^{(t)}(a_{1:N}^{sy}, a_{0:N}^{su})$.*

Theorem 6.2 *If $J_{EN}(a^{sy}, a^{su})$ in* Problem 3.3 *is coercive in $a_{1:N}^{sy}$ and $a_{0:N}^{su}$, then $J_{EN}^{(t)}(a_{1:N}^{sy}, a_{0:N}^{su})$ is coercive in $a_{1:N}^{sy}$ and $a_{0:N}^{su}$ for sufficiently large t. In this case, if $a_{1:N}^{sy}$ and $a_{0:N}^{su}$ solve* Problem 3.3 *and $a_{1:N}^{sy}(t)$ and $a_{0:N}^{su}(t)$ solve* Problem 6.1, *then*

$$\lim_{t\to\infty} a_{1:N}^{sy}(t) = a_{1:N}^{sy}, \qquad \lim_{t\to\infty} a_{0:N}^{su}(t) = a_{0:N}^{su}. \tag{6.3}$$

The proof of Theorem 6.2 is a special case of the proof of Theorem 6.5.

Because of the finite dimensionality assumed in this section, all norms for $a_{1:N}^{sy}$ and $a_{0:N}^{su}$ are equivalent, so that the limits in (6.3) and all other limits in this section are unambiguous.

6.2 Minimax Estimation and Prediction

For each $t \geq 0$, we define $J_{EN}^{(t)}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz})$ by the right-hand side of (6.1) with

$$\varepsilon_k(t) \;=\; \sum_{i=0}^{N}\langle u_{k-i}, a_i^{su}\rangle + \sum_{i=1}^{N}\langle y_{k-i}, a_i^{sy}\rangle + \sum_{i=1}^{N}\langle z_{k-i}, a_i^{sz}\rangle \;-\; s_k, \qquad 0 \leq k \leq t, \tag{6.4}$$

with $y_i = 0$, $z_i = 0$, and $u_i = 0$ for $i < 0$. Also, we define $\tilde{J}_{EN}^{(t)}(a_{1:N}^{sy}, a_{1:N}^{sz})$ as in (4.38) with $J_{EN}(a_{1:N}^{sy}, a_{0:N}, a_{1:N}^{sz})$ replaced by $J_{EN}^{(t)}(a_{1:N}^{sy}, a_{0:N}, a_{1:N}^{sz})$.

Problem 6.3 Problem 4.3 *with $J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz})$ replaced by $J_{EN}^{(t)}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz})$.*

Problem 6.4 Problem 4.12 *with $\tilde{J}_{EN}(a_{1:N}^{sy}, a_{1:N}^{sz})$ replaced by $\tilde{J}_{EN}^{(t)}(a_{1:N}^{sy}, a_{1:N}^{sz})$ and $J_{EN}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz})$ replaced by $J_{EN}^{(t)}(a_{1:N}^{sy}, a_{0:N}^{su}, a_{1:N}^{sz})$.*

Theorem 6.5 *If the coercivity hypotheses in* Problem 4.3 *hold, then the coercivity hypotheses in* Problem 6.3 *hold for sufficiently large t. In this case, if $a_{1:N}^{sy}$, $a_{0:N}^{su}$, and $a_{1:N}^{sz}$ solve* Problem 4.3 *and $a_{1:N}^{sy}(t)$, $a_{0:N}^{su}(t)$, and $a_{1:N}^{sz}(t)$ solve* Problem 6.3, *then*

$$\lim_{t\to\infty} a_{1:N}^{sy}(t) = a_{1:N}^{sy}, \qquad \lim_{t\to\infty} a_{0:N}^{su}(t) = a_{0:N}^{su}, \qquad \lim_{t\to\infty} a_{1:N}^{sz}(t) = a_{1:N}^{sz}. \tag{6.5}$$

Similarly for Problems 4.12 *and* 6.4.

Proof Problems 4.3 and 6.3: Since U, Y, and Z are finite-dimensional, $a^{su}_{0:N}$, $a^{sy}_{1:N}$, and $a^{sz}_{1:N}$ can be represented, respectively, by finite-dimensional column vectors α^u, α^y, and α^z, all of which can be included in a parameter vector $\alpha \in \mathbf{C}^n$, where n is the total number of scalar parameters in $a^{su}_{0:N}$, $a^{sy}_{1:N}$, and $a^{sz}_{1:N}$. Then

$$\begin{aligned} J_{EN}(a^{sy}_{1:N}, a^{su}_{0:N}, a^{sz}_{1:N}) &= \langle Q\alpha, \alpha\rangle + 2\Re e\langle \beta, \alpha\rangle + J_{EN}(0,0,0), \\ J^{(t)}_{EN}(a^{sy}_{1:N}, a^{su}_{0:N}, a^{sz}_{1:N}) &= \langle Q(t)\alpha, \alpha\rangle + 2\Re e\langle \beta(t), \alpha\rangle + J^{(t)}_{EN}(0,0,0), \end{aligned} \tag{6.6}$$

where $\beta, \beta(t) \in \mathbf{C}^n$, Q and $Q(t)$ are $n \times n$ matrices and $\langle\cdot,\cdot\rangle$ is the Euclidean inner product on $\mathbf{C}^n$. The coercivity hypotheses in the parameter-estimation problems mean that certain submatrices of Q and $Q(t)$ must be either positive definite or negative definite. When these conditions hold, Q and $Q(t)$ are nonsingular and the solutions to Problems 4.3 and 6.3 are (for sufficiently large t)

$$\begin{aligned} \alpha &= -Q^{-1}\beta, \\ \alpha(t) &= -Q^{-1}(t)\beta(t). \end{aligned} \tag{6.7}$$

It follows from the definitions in Section 2, the operator definitions at the beginning of Section 3, and the definitions of $J_{EN}(a^{sy}_{1:N}, a^{su}_{0:N}, a^{sz}_{1:N})$ and $J^{(t)}_{EN}(a^{sy}_{1:N}, a^{su}_{0:N}, a^{sz}_{1:N})$ that the elements of Q and β are the limits of the corresponding elements of $Q(t)$ and $\beta(t)$. Therefore, when Q is nonsingular, $\lim_{t\to\infty} Q^{-1}(t) = Q^{-1}$ and $\lim_{t\to\infty}\alpha(t) = \alpha$.

Essentially the same argument works for Problems 4.12 and 6.4. □

For $1 \le k \le N$, we define $J^{(t)}_{EN,k}(a^{sy}_{1:N}, a^{su}_{0:N}, a^{sz}_{1:N})$ as in (4.8)–(4.10) but with J_{EN} replaced by $J^{(t)}_{EN}$ in (4.8). Also, we define $\tilde{J}^{(t)}_{EN,k}(a^{sy}_{1:N}, a^{sz}_{1:N})$ as in (4.38) and (4.41) with J_{EN} replaced by $J^{(t)}_{EN}$ in (4.38). Theorems 4.4 and 4.13 hold with $J_{EN,k}$ and $\tilde{J}_{EN,k}$ replaced by $J^{(t)}_{EN,k}$ and $\tilde{J}^{(t)}_{EN,k}$, respectively.

Problem 6.6 Problem 4.5 *with* $J_{EN,k}(a^{sy}_{1:N}, a^{su}_{0:N}, a^{sz}_{1:N})$ *replaced by* $J^{(t)}_{EN,k}(a^{sy}_{1:N}, a^{su}_{0:N}, a^{sz}_{1:N})$ *in* Theorem 4.4 *and* Problem 4.5.

Problem 6.7 Problem 4.14 *with* $\tilde{J}_{EN,k}(a^{sy}_{1:N}, a^{sz}_{1:N})$ *replaced by* $\tilde{J}^{(t)}_{EN,k}(a^{sy}_{1:N}, a^{sz}_{1:N})$ *in* Theorem 4.13 *and* Problem 4.14 *and* $J_{EN}(a^{sy}_{1:N}, a^{su}_{0:N}, 0)$ *replaced by* $J^{(t)}_{EN}(a^{sy}_{1:N}, a^{su}_{0:N}, 0)$.

Theorem 6.8 *If the coercivity hypotheses in* Problem 4.5 *hold, then the coercivity hypotheses in* Problem 6.6 *hold for sufficiently large t. In this case, if $a^{sy}_{1:N}$ and $a^{su}_{0:N}$ solve* Problem 4.5 *and $a^{sy}_{1:N}(t)$ and $a^{su}_{0:N}(t)$ solve* Problem 6.6, *then*

$$\lim_{t\to\infty} a^{sy}_{1:N}(t) = a^{sy}_{1:N}, \qquad \lim_{t\to\infty} a^{su}_{0:N}(t) = a^{su}_{0:N}. \tag{6.8}$$

Similarly for Problems 4.14 *and* 6.7.

Proof Problems 4.5 and 6.6: For $1 \le k \le N$, all of the parameters in $a^{sy}_{1:k}$, $a^{su}_{0:k}$, and $a^{sz}_{1:k}$ can be represented by a column vector α_k, and

$$\begin{aligned} J_{EN,k}(a^{sy}_{1:k}, a^{su}_{0:k}, a^{sz}_{1:k}) &= \langle Q_k\alpha_k, \alpha_k\rangle + 2\Re e\langle \beta_k, \alpha_k\rangle + J_{EN,k}(0,0,0), \\ J^{(t)}_{EN,k}(a^{sy}_{1:k}, a^{su}_{0:k}, a^{sz}_{1:k}) &= \langle Q_k(t)\alpha_k, \alpha_k\rangle + 2\Re e\langle \beta_k(t), \alpha_k\rangle + J^{(t)}_{EN,k}(0,0,0). \end{aligned} \tag{6.9}$$

The matrix Q_k and the column vector β_k are obtained by a finite number of matrix multiplications, additions, and inversions starting with submatrices of the Q and β in (6.6), and similarly for $Q_k(t)$ and $\beta_k(t)$. Hence $Q_k(t)$ and $\beta_k(t)$ are continuous functions of $Q(t)$ and $\beta(t)$, so that $Q_k(t)$ and $\beta_k(t)$ converge to Q_k and β_k as $Q(t)$ and $\beta(t)$ converge to Q and β. It follows that, for each k, the solution to the linear-quadratice minimax problem on the right-hand side of (4.9), with $J_{EN,k+1}$ replaced by $J^{(t)}_{EN,k+1}$, converges to the solution of the original minimax problem on the right-hand side of (4.9).

Essentially the same argument works for Problems 4.14 and 6.7. □

7 Conclusions

From the way we use norms and inner products of equivalence classes of scalar sequences in defining and analyzing the parameter-estimation problems, it follows that solutions to any of the parameter-estimation problems in Sections 3 and 4 and the asymptotic values of the estimated parameters in any of the problems in Section 6 depend only on the equivalence classes to which u, w, s, y, and z belong. Therefore, the asymptotic parameter estimates are unaffected by changing a finite number of terms in any of the measurement sequences or the known input sequence u or by changing the initial state vector of the plant, since the plant is assumed to be exponentially stable. This is to be expected, and the fact that it is an immediate consequence of our formulation suggests that Hilbert spaces containing equivalence classes of deterministic scalar sequences provide a natural setting for many parameter-estimation problems.

The class of least-squares problems in Section 3 includes many common problems in parameter estimation and adaptive filtering and prediction, but to keep the discussion manageable, we have omitted several classes of problems that can be analyzed by straightforward generalizations of the methods here. For example, the results in Sections 3–6 can be generalized to problems with correlated process and sensor noise and problems in which the summations over the y terms and u terms in the ARX model have different limits N_y and N_u instead of the same limit N.

The results in Sections 3–5 characterize fully the asymptotic values of least-squares or minimax parameter estimates, but there are only two possibilities for computing these limiting values for a particular problem: (1) solution of the parameter-estimation problem for long input/output data sequences, (2) solution of either the corresponding state-space control problem or the corresponding state-space filtering problem, either of which requires complete knowledge of the plant and noise statistics. Of course, if the information required for (2) is known, then there is no reason for parameter estimation.

On the other hand, it often occurs that some rough estimates of plant and noise characteristics are known but the *a priori* information is not sufficiently accurate for designing a filter or perhaps a controller. In this case, the characterizations in this paper can be used to compute rough indications of the results that will be obtained with the different parameter-estimation problems and for different ARX orders. For example, comparing Markov parameters for steady-state and finite-interval Kalman filters computed for *a priori* models of the plant and noise should indicate how large a finite-dimensional ARX model must be used in LS identification if the estimated parameters are to approximate the Markov parameters of the steady-state Kalman filter for the true plant and noise. Such *a priori* comparison should be useful in applying the OKID method for system identification [3, 4, 5]. With *a priori* plant and noise models, (3.27) and (3.28) can be used to compare roughly the performance levels to be expected of adaptive LS filters of different orders, since the optimal values of the performance indices in the control problems are easy to compute with Riccati matrices. Similar *a priori* comparisons between LS and minimax parameter estimation and filtering can be obtained with (3.37) and (3.41). Also, the control problems in Section 4 can be used for *a priori* models of the plant and noise to get rough *a priori* lower bounds for the values of γ that can used in minimax parameter estimation and filtering for the true plant.

References

[1] A.V. Balakrishnan, *Applied Functional Analysis,* Second Edition. New York: Springer-Verlag, 1981.

[2] R.F. Curtain and A.J. Pritchard, *Infinite Dimensional Linear Systems Theory.* New York: Springer-Verlag, 1978.

[3] C.-W. Chen, J.-K. Huang, M. Phan, and J.-N. Juang, "Integrated system identification and state estimation for control of flexible space structures," *Journal of Guidance, Control, and Dynamics*, vol. 15, pp. 88–95, January–February 1992.

[4] J.-N. Juang and R. S. Pappa, "Effects of noises on modal parameters identified by the eigensystem realization algorithm," *Journal of Guidance Control and Dynamics*, vol. 9, pp. 294–303, May-June 1986.

[5] J.-N. Juang, M. Phan, L.G. Horta, and R.W. Longman, "Identification of observer/Kalman filter Markov parameters: theory and experiments," in *AIAA Guidance Navigation and Control Conference*, (New Orleans, LA), pp. 1195–1207, AIAA, August 1991.

[6] T. Basar and P. Bernhard, *H^∞-Optimal control and related minimax design problems.* Boston: Birkhauser, 1991.

[7] K.M. Nagpal and P.P. Khargonekar, "Filtering and smoothing in an H^∞ setting," *IEEE Transactions on Automatic Control*, vol. 36, pp. 152–166, February 1991.

[8] U. Shaked and Y. Theodor, "A frequency domain approach to the problems of H_∞-minimum error state estimation and deconvolution," *IEEE Transactions on Signal Processing*, vol. 40, pp. 3001–3011, December 1992.

[9] U. Shaked, "H_∞-minimum error state estimation of linear stationary processes," *IEEE Transactions on Automatic Control*, vol. 35, pp. 554–558, May 1990.

[10] I. Yaesh and U. Shaked, "A transfer function approach to the problems of discrete-time systems: H_∞ optimal linear control and filtering," *IEEE Transactions on Automatic Control*, vol. 36, pp. 1264–1271, November 1991.

[11] U. Shaked and Y. Theodor, "H_∞ optimal estimation: a tutorial," in *31st Conference on Decision and Control*, (Tucson, AZ), pp. 2278–2286, IEEE, December 1992.

[12] J.L. Speyer, C.-H. Fan, and R.N. Banavar, "Optimal stochastic estimation with exponential cost criteria," in *31th Conference on Decision and Control*, (Tucson, AZ), pp. 2293–2298, IEEE, December 1992.

[13] P. P. Khargonekar and M. A. Rotea, "Mixed H_2/H_∞ filtering," in *31st Conference on Decision and Control*, (Tucson, AZ), pp. 2299–2304, IEEE, December 1992.

[14] M. J. Grimble, R. Hashim, and U. Shaked, "Identification algorithms based on H_∞ state-space filtering techniques," in *31th Conference on Decision and Control*, (Tucson, AZ), pp. 2287–2292, IEEE, December 1992.

[15] I. Yaesh and U. Shaked, "Game theory approach to optimal linear estimation in the minimum H_∞ norm sense," in *28th Conference on Decision and Control*, (Tampa, FL), pp. 421–425, IEEE, December 1989.

[16] I. Yaesh and U. Shaked, "H_∞ optimal estimation—the discrete time case," in *9th International Symposium on MTNS*, (Kobe, Japan), pp. 261–267, June 1991.

[17] L. Xie, C. E. De Souza, and M. Fu, "H_∞ estimation for discrete-time linear uncertain systems," *International Journal of Robust and Nonlinear Control*, vol. 1, pp. 111–123, 1991.

[18] M.J. Grimble, "Polynomial matrix solution of the H_∞ filtering problem and the relationship to Riccati equation state-space results," *IEEE Transactions on Signal Processing*, pp. 67–81, January 1993.

[19] L. Ljung, *System Identification: Theory for The User*. Englewood Cliffs, New Jersey: Prentice-Hall, 1987.

[20] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*. Englewood Cliffs, New Jersey: Prentice-Hall, 1979.